

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 00:05:35 ; Search time 24.93 seconds
(without alignments)
738.744 Million cell updates/sec

Title: US-09-837-602-2
Perfect score: 3899
Sequence: 1 MWKLLPAGGAGGEPYRLLT.....KEESLADDFRYNPLYKRRR 754

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------------|--------------------|
| 1 | 170.5 | 4.4 | 1939 | US-09-310-187A-1 | Sequence 1, Appli |
| 2 | 169 | 4.3 | 1128 | US-08-923-992A-6 | Sequence 6, Appli |
| 3 | 165 | 4.2 | 1886 | US-08-938-105-3 | Sequence 3, Appli |
| 4 | 161 | 4.1 | 1098 | US-08-923-992A-8 | Sequence 8, Appli |
| 5 | 158.5 | 4.1 | 1664 | US-08-642-846-2 | Sequence 2, Appli |
| 6 | 158.5 | 4.1 | 1664 | US-09-264-604-2 | Sequence 2, Appli |
| 7 | 157.5 | 4.0 | 976 | US-09-104-324B-4 | Sequence 4, Appli |
| 8 | 157 | 4.0 | 1164 | US-08-923-992A-2 | Sequence 2, Appli |
| 9 | 156 | 4.0 | 1164 | US-08-923-992A-10 | Sequence 10, Appli |
| 10 | 152.5 | 3.9 | 754 | US-09-214-564A-2 | Sequence 2, Appli |
| 11 | 150 | 3.8 | 461 | US-08-630-822A-68 | Sequence 68, Appli |
| 12 | 150 | 3.8 | 461 | US-09-005-069-68 | Sequence 68, Appli |
| 13 | 146.5 | 3.8 | 1093 | PCT-US93-03077-1 | Sequence 1, Appli |
| 14 | 146.5 | 3.8 | 3248 | US-08-353-700-1 | Sequence 1, Appli |
| 15 | 146.5 | 3.8 | 3248 | PCT-US95-16216-1 | Sequence 1, Appli |
| 16 | 145 | 3.7 | 1702 | US-08-296-791-5 | Sequence 5, Appli |
| 17 | 145 | 3.7 | 1702 | PCT-US95-10661A-5 | Sequence 5, Appli |
| 18 | 144.5 | 3.7 | 1014 | US-09-078-347A-3 | Sequence 3, Appli |
| 19 | 144 | 3.7 | 1588 | PCT-US93-07261-11 | Sequence 11, Appli |
| 20 | 144 | 3.7 | 1663 | PCT-US93-07261-16 | Sequence 16, Appli |
| 21 | 143 | 3.7 | 904 | US-09-198-484-2 | Sequence 2, Appli |
| 22 | 143 | 3.7 | 1013 | US-08-860-886-2 | Sequence 2, Appli |
| 23 | 142.5 | 3.7 | 2482 | US-08-328-254-6 | Sequence 6, Appli |
| 24 | 141.5 | 3.6 | 1120 | US-09-147-404-1 | Sequence 1, Appli |
| 25 | 141.5 | 3.6 | 1354 | US-08-685-871-2 | Sequence 2, Appli |
| 26 | 141 | 3.6 | 1494 | US-08-755-587-186 | Sequence 186, App |
| 27 | 140 | 3.6 | 1435 | US-08-568-459A-4 | Sequence 4, Appli |

| | | | | | | |
|----|-------|-----|------|---|-------------------|--------------------|
| 28 | 140 | 3.6 | 1435 | 2 | US-08-487-826B-4 | Sequence 4, Appli |
| 29 | 139.5 | 3.6 | 984 | 1 | US-08-242-932-2 | Sequence 2, Appli |
| 30 | 139.5 | 3.6 | 984 | 1 | US-08-714-481-2 | Sequence 2, Appli |
| 31 | 139.5 | 3.6 | 984 | 5 | PCT-US95-06111-2 | Sequence 2, Appli |
| 32 | 139 | 3.6 | 1104 | 4 | US-08-923-992A-4 | Sequence 4, Appli |
| 33 | 138.5 | 3.6 | 1786 | 4 | US-08-973-462-8 | Sequence 8, Appli |
| 34 | 136.5 | 3.5 | 3788 | 4 | US-09-336-447A-76 | Sequence 76, Appli |
| 35 | 136 | 3.5 | 1075 | 2 | US-08-993-228-19 | Sequence 19, Appli |
| 36 | 135.5 | 3.5 | 683 | 6 | 5210183-3 | Patent No. 5210183 |
| 37 | 135.5 | 3.5 | 1201 | 4 | US-09-098-901-2 | Sequence 2, Appli |
| 38 | 135 | 3.5 | 1036 | 4 | US-08-891-640-3 | Sequence 3, Appli |
| 39 | 134 | 3.4 | 1184 | 4 | US-09-541-782-2 | Sequence 2, Appli |
| 40 | 134 | 3.4 | 2285 | 4 | US-09-308-375-2 | Sequence 2, Appli |
| 41 | 133 | 3.4 | 376 | 6 | 5180810-1 | Patent No. 5180810 |
| 42 | 131.5 | 3.4 | 558 | 1 | US-08-285-440-6 | Sequence 6, Appli |
| 43 | 131.5 | 3.4 | 558 | 1 | US-08-630-349-6 | Sequence 6, Appli |
| 44 | 131.5 | 3.4 | 2343 | 4 | US-09-324-867-2 | Sequence 2, Appli |
| 45 | 131 | 3.4 | 595 | 1 | US-08-171-718-16 | Sequence 16, Appli |

ALIGNMENTS

RESULT 1
US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751
; GENERAL INFORMATION:
; APPLICANT: Benichou, Gilles
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; TITLE OF INVENTION: Graft Rejection
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310,187A
; CURRENT FILING DATE: 1999-05-12
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-310-187A-1

| | | | | |
|-----------------------|-------|--|-------|--------------------------------------|
| Query Match | 4.4% | Score 170.5; | DB 4; | Length 1939; |
| Best Local Similarity | 23.4% | Pred. No. 8 6e-05; | | |
| Matches | 91; | Conservative | 57; | Mismatches 132; Indels 109; Gaps 17; |
| QY | 388 | KMPQKFRMLSDQAPTQVKSCKTSSNNNSMVSNTLAKMRIPNYOLSPKLSINKSDRAS | 447 | |
| DB | 833 | KLYFKIKPLKLSAETKEMA-TMKEEFGRIKETLEKSEARRKELE-EKMYSLQLQEKNDLQ | 890 | |
| QY | 448 | QQOQTS-----IRNYFQFSTKKR---ERDEENQEMS---SCKSARIETSCS | 488 | |
| DB | 891 | LQVQAEQDNLNDAERCDQLIKNKIQLEAKVKEMNERLEDEENNAELTAKRKLEDCS | 950 | |
| QY | 489 | -LLEQTOPATPPLSWK-NKEOHLSENPEVDNTSDNNLFTDRLSKSVKNSKSHAAEKL | 546 | |
| DB | 951 | ELAKDDIDDLTLAKVEKEKHATENK-----VKNLTEMAGLDEIIAKLT | 995 | |
| QY | 547 | SNKK-----REMDVDAIDEVLEQLFKDTPKELEIDVKVQOEEDVNYKRRPMDIET | 599 | |
| DB | 996 | KEKKALQEAHQALDDLDQVEEDKVNSL---SKSKVLEQQVDDLEGSLEQEKVYMDLER | 1052 | |
| QY | 600 | NTFSDAEVPESSKISQENFEGKRELKEDSLWSAKEISN--NDKLODDSEMLPKKLLLT | 657 | |
| DB | 1053 | -----AKRLEGDLKLTQESIMDLNDKLO-----LEEKLKKK | 1085 | |
| QY | 658 | EFPSLYIKNSTRNPSGINDDYG-QLKNFKKKFKVTPYGAGKPLHIITGSDLIHAHAKN | 716 | |
| DB | 1086 | EF-----DINQNSKIEDEQALQLOKRLKE-----NOARI | 1117 | |
| QY | 717 | TELEWLROEMEVQNOHAKESLADDLFR | 745 | |

```
Query Match      4.3%; Score 169; DB 4; Length 1128;
Best Local Similarity 19.9%; Pred. No. 5.1e-05;
Matches 142; Conservative 125; Mismatches 238; Indels 208; Gaps 38;
```

| | |
|----|--|
| QY | 192 QIESFVPPDPEPSIGSKNVDLSGR-----OERKQIFPKGTFITFLNAQHKKLUSSA 241 : : : : : : : : : : : : : : : : : |
| Dd | 95 QKNFEFTKIDETDSDALLEQNQFNETRNLHIIKHOOEEVEKD----RAKOQTTLQS 149 : : : : : : : : : : : : : : : : : |
| QY | 242 VFGGGEARLIITEENEHEHNFFLAPGTCCVD--TCITSNSTQLIPDCOKKWITQSIDMMLQRQ 300 : : : : : : : : : : : : : : : : : |
| Dd | 150 -----DTKVLDLNIDKLHQNSPVEKMAEPKGITNED---KDSMLUKKIEDIRKQAQA 200 : : : : : : : : : : : : : : : : : |
| QY | 301 GLRPIPEAEI--GLAVFMFTTKNYCDPGQHPSTGLKTTPGPSLSQGVSDEKLMPSPV 358 : : : : : : : : : : : : : : : : : |
| Dd | 201 DKREDAEVKYREEBLKLFSSTKAGLOEIEIHVKKEFS-----SEENTQKVDEHY-----A 251 : : : : : : : : : : : : : : : : : |
| QY | 359 NTTYTVADTSESQADTWDLSERPKEIKVSKME--QK-----FRMLSODAP 401 : : : : : : : : : : : : : : : : : |
| Dd | 252 NSLONLAQKSLEELDRATTNEQATOVKNQFOLENAAQKLKEMQPLIKETNVLYKAMSESLE 311 : : : : : : : : : : : : : : : : : |
| QY | 402 TVRESCKTSNNN--SMVSNFTLAMRIPTVOLSPTK--LPINSINKSDBASO--QOOTNSIR 456 : : : : : : : : : : : : : : : : : |
| Dd | 312 QBEBELKHNSEANLEDLVAKSAISIVREYGKINQSLKNLPOLKOLEEAHSKLGKVOEDFR 371 : : : : : : : : : : : : : : : : : |

Query Match 4.2%; Score 165; DB 4; Length 1886;
Best Local Similarity 22.5%; Pred. No. 0.00023;
Matches 91; Conservative 61; Mismatches 123; Indels 13

Gaps 17;


```

> COUNTRY: USA
> ZIP: 55401
> COMPUTER READABLE FORM:
> MEDIUM TYPE: Floppy disk
> COMPUTER: IBM PC compatible
> OPERATING SYSTEM: PC-DOS/MS-DOS
> SOFTWARE: PatentIn Release #1.0, Version #1.30
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/642.846
> FILING DATE: 03-MAY-1996
> CLASSIFICATION: 424
> ATTORNEY/AGENT INFORMATION:
> NAME: MUETING, ANN M.
> REGISTRATION NUMBER: 33,977
> REFERENCE/DOCKET NUMBER: 110.00280101
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: 612-305-1217
> TELEFAX: 612-305-1228
> INFORMATION FOR SEQ ID NO: 2:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 1664 amino acids
> TYPE: amino acid
> STRANDEDNESS: single
> TOPOLOGY: linear
> MOLECULE TYPE: protein
> US-08-642-846-2

```

Query Match 4.1%; Score 158.5; DB 2; Length 1664;
Best Local Similarity 21.0%; Pred. No. 0.00063;
Matches 105; Conservative 74; Mismatches 216; Indels 105; Gaps 20;

| | | | |
|----|-----|--|------|
| QY | 257 | EEBHNFFLAPGTCVDTGTNSOTFLPDCQKWKIQSIMDMLOQOGLURPPEAEIGLAVIF | 3116 |
| Db | 272 | DERKNVLYSLP-----TNNNNSKN-YSDMDSH-LQNLQDASKNKNENTHNLFSALKAPK | 323 |
| QY | 317 | MTTKNTCDQPGHPSTGLTKTTTPGSLSGVSGVDEKLMPSAP-----VNTTTVVADETSQ | 371 |
| Db | 324 | NDIENPLNSLTNADISILRSSGSSQSSLOSLRNDNRVLESVPGSPKKVNPGLSLNDGKGF | 383 |
| QY | 372 | ADTWDLSERPKEIKVSKMEQKFRMLSQDAPTVKESCCTSTNNNSMVSNTLAKMRIPNY-Q | 430 |
| Db | 384 | SDEVESLLPRLDSROKLE-----TTKEHDAPEHNNEFID---AKSTNTNKGQ | 429 |
| QY | 431 | L-----SPTKLPSINKSDRASQQOQTNSIRNYFQPTKKR-----ERDEENQEMSSKSA | 481 |
| Db | 430 | LLVSSDDHLDSFDRSYNHTQ-----SILNLLNSASQOISLNALEKQRTQEQ----- | 478 |
| QY | 482 | RIETSCSLLEQOTOPATPSLWKNKEQHLSENPVDYNSDNLLTDDTLKISIVKNSASKSHA | 541 |
| Db | 479 | -----EQTQAAEP-----EETSFSQNIKVKQEPKSNL-----EFVFKTITKEPV | 518 |
| QY | 542 | AEKLRNKKKREMD-----VAIEDEVLEQLFKDTKPELIDVKVQOEVDNVRKPRMDI | 597 |
| Db | 519 | SATEIKAPREFSSRILIRKNEDEIAEP-ADLHPKKEANEASHVEDTDLALKK----- | 570 |
| QY | 598 | ETWDTFSDNAVPESSKISQENETGKARELKEDSLWSAKEISNNDKLQDDSEMLPKKLLT | 657 |
| Db | 571 | ALRNDDEESTQTQNSTKMSIRFHI-----DSDMKLEDSNDGDREDND-----IS | 614 |
| QY | 658 | EFRSLVTKNSTSRNPSSGINDVYGOLKNFKFKVTVPGAGKLPHIIGGSDLIAHARKNT | 717 |
| Db | 615 | RFEKSDILNDVQSOTSDIGDKYGNSSSEITTKTLAPRS-----DNDKNKSNKSLDP | 667 |
| QY | 718 | ELEEWLRQEMEVONQHAKEE | 737 |
| Db | 668 | ANNESLQQQLEEV---PHTKD | 685 |

RESULT 6
US-09-364-604-2
; Sequence 2, Application US/09264604
; Patent No. 6346411

; GENERAL INFORMATION:
 ; APPLICANT: HOSTETTER, MARGARET K.
 ; APPLICANT: GALE, CHERYL A.
 ; APPLICANT: BENDEL, CATHERINE M.
 ; APPLICANT: TAO, NIAN-JUN
 ; APPLICANT: KENDRICK, KATHLEEN
 ; TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE,
 ; TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MUETING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
 ; STREET: 119 NORTH FOURTH STREET, SUITE 203
 ; CITY: MINNEAPOLIS
 ; STATE: MINNESOTA
 ; COUNTRY: USA
 ; ZIP: 55401
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/264,604
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/642,846
 ; FILING DATE: 03-MAY-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: MUETING, ANN M.
 ; REGISTRATION NUMBER: 33,977
 ; REFERENCE/DOCKET NUMBER: 110.00280101
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 612-305-1217
 ; TELEFAX: 612-305-1228
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1664 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-264-604-2

| | | | | |
|---------------------------|--------|--------------------|-------------|--------------|
| Query Match | 4.1% | Score 158.5; | DB 4; | Length 1664; |
| Best Local Similarity | 21.0%; | Pred. No. 0.00063; | | |
| Matches 105: Conservative | 74; | Mismatches 216; | Indels 105; | Gaps 20; |

| | | | |
|----|-----|---|-----|
| Qy | 257 | E E B H N F L A P G T C V D T G I T N S O T L I P D C Q K W I Q S I M D M L R Q R U R P I P E A E I G L A V I F | 316 |
| | : | : : : | : |
| Db | 272 | D E X V N Y L S P - - - - - T N N S K N - V S D M D S H - L Q N L O A S K N K T N E I N H L S F A L K A P K | 323 |
| Qy | 317 | M T T K N Y C D P G H P S T G L K T T P G P S I S Q G V S V D E K L M P S A P - - - - - V N T T T V A V D E S Q | 371 |
| | : | : : : | : |
| Db | 324 | N D I E N P L N S I T N A D I L S R S G S Q S Q L S R J D N D N R V L E S V P G S P K V N P G L S I N D I K G F | 383 |
| Qy | 372 | A D T W D L S E R P E K I K V S M E Q K F M L S Q D A P T V K E S C K T S N N N S M Y S N T L A K M R I P N Y - Q | 430 |
| | : | : : : | : |
| Db | 384 | S D E V S L D P R J S R O K L E - - - - - T T K E H D A P E H N N E F I D - - - A K S T N T K G Q | 429 |
| Qy | 431 | L - - - S P T K L P S I N K S D R A S Q O O T N S I R N Y F O P S T K K R - - - - - E R D E N O E M S S K S A | 481 |
| | : | : : : | : |
| Db | 430 | L V S S D D H L D S P R S Y N H T E Q - - - - - S I L N L N S A S Q S I S L N A L E K Q R O T O E Q - - - - - | 478 |
| Qy | 482 | R I E T S C L L B Q T O P A T P S L W K N K E Q H L S E N P Y D T N S D N L T D T D L K S I V K N S A S K S H A | 541 |
| | : | : : : | : |
| Db | 479 | - - - - - E Q T Q A E P - - - - - E E T S F S D N I K V Q E P K S N L - - - - - E F V K V T I K K E P V | 518 |
| Qy | 542 | A E K L R N K K R E M D - - - - - V A I E D V E L U F K P D K P E L E I D V K V Q O E D V N V K R P R M D I | 597 |
| | : | : : : | : |
| Db | 519 | S A T E I A K P R E F S S I R L I R K N E D E I A E P - - A D I H P K K E A N E A S H V B D T D A L L A K - - - - - | 570 |

QY 598 ETNDSDEAVPSSKISQENEIGKGRKELKEDSLWSAKEISNNNDKLODDSEMLPKKLLLT 657
Db 571 ALNDDESDTTQNTKMSIRFHI-----DSDWLEDSNDGDRDND-----IS 614
QY 658 EPRSLLVKNSTSRNPSSINDDYQOLKNFKKVTYPGACKLPHIIGGSDLIHAKHAKNT 717
Db 615 RPKSDILNDVSTSDIIGDKYGNSSSEITTKTLAPRS-----DNNDKENSKSLEDP 667
QY 718 ELEEWLRQEMEVQONHAKKE 737
Db 668 ANNESLQOQLEV--PHTKED 685
RESULT 7
US-09-104-324B-4
; Sequence 4, Application US/09104324B
; Patent No. 6232460
; GENERAL INFORMATION:
; APPLICANT: T reci, Ozlem; Sahin, Ugur; Pfreundschuh, Michael
; TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,
; TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample Of
; TITLE OF INVENTION: No. 6232460mal Cells
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski LLP
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,324B
; FILING DATE: 25-June-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/892,702
; FILING DATE: 15-July-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6232460man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5491
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 318-3000
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-09-104-324B-4
Query Match 4.0%; Score 157.5; DB 4; Length 976;
Best Local Similarity 19.0%; Pred. No. 0.00035;
Matches 159; Conservative 150; Mismatches 312; Indels 215; Gaps 41;
QY 22 VEYVGRKNCALIIENDQSIQRHVAFLTANFSVTNLSQTDIEPVLTKDONS-----K 73
Db 46 LEPPFAKTN---LSKNGENIDSPALQKVN-----LPVLEQVGNSDCHYQEGKL 92
QY 74 YGTFVNEEKMQNGFSRTLSGDIITGVGSKFRIEYE-PLVACSSCLDVSGK--TALNQ 130
Db 93 DSDLENSGLSRVFSKLYKEAKI-----KKWVSTEAELRQESKQLQENRRIIEAQRK 146
QY 131 AI--LQLGGFTVNNWTEE--CTHLVMSVVKVTIKTICALI---CGRPVVKP-----EYFTEF 180
Db 147 AIQELQFGNEKVSLEEGIOENKDLIKENNATRLCNLLKETCARSAEKTKKYEVEREE 206
QY 181 LKAVQSKKQPPQIESPYPLDEPSIGSKN-----VDLSGROERKQIFKGKTFIFL 230

Db 207 TRGVYMDLN--NNIEKMITAFQELRVQAENSRLEMHFKLDEYKIQHLEQYKKE-----I 261
QY 231 NAKQHKLLSSAVVFGGGEARLITEENEEHFFFLAPCTCVVDGIGTNSQTLIPDCOKKWI 290
Db 262 NDKE-KQVSLLLI-----QITEKENMKMDLTFL-----LEESRQKVQOLEK-- 302
QY 291 QSIMDMLQROGLRPIPEAEIGLAVIFMTTKNYCDPOGHPSTGLKTTTPGSLSQGVSVDE 350
Db 303 -----TKLQSENL-----KQSTIEKHHLTKLELDIK--VSLQRSVSTQK 339
QY 351 KLMPSPAPVNTTYVADTSEQ-----ADTWDLSE-----RPKEIKVS 387
Db 340 ALBEDLIQATKTKICQTEKEKQMETESNKARAAHSFVVFETVTVCSLEELLRTQOORLE 399
QY 388 KMEQKFRMLSOD----APTVKESCKTSSNNN---SMVSNFTLAKMRIPNY-----Q 430
Db 400 KNEQDKILTMELQKKSSSELEEMTKLTNNKEVELEELKKVLGKEKTELLYENKQFEKTAEE 459
QY 431 LSPTKLPSIN-----KSKDRASQOQOQNSI---RNYFQPSKKRERDEENQEM-----SS 477
Db 460 LKQTEQELIGLLQAREKEVHDLEITQLTAITTSEQYYSKEVKDLKTELENEKLNKLTSLH 519
QY 478 CKSARIETSCSLEQTPATPSLWKNKEQHLSENPEVDNNS-----DNNLFTDLDLSIVK 533
Db 520 CNKLSLENK-ELTQETSDMTLEL-KNOQEDINNKKOEERMLKQIENLQETETQLRNELE 577
QY 534 -----NSAKSHAAEKLSRKNKREMDVATEDEVLEQLFKDKTKPELE 575
Db 578 YVREELKQKRDYVKCKLDKSEENCNLRKOVENKNKYIEELQOENKALKKGTAESKQLN 637
QY 576 I-DVKVQKQEDYVNRKPRMDIETNDTFSDEAVPSSKISQEN---EIGKKRELKEDSL 631
Db 638 VYEIKVKNLELESQAKQFGEI--TDTYQKEI--EDKKISEENLLLEEVEKAKVIADEAV 693
QY 632 WSAKEISN--NDKLQDDSEMLPK-----KLLITEPRSLVKKTSRNPSSGINDDYG-QL 682
Db 694 KLOKEIDKRCQHKIAEMVAMLEKHKHQYDKIIEERSELGLYKSKEOQSLSRASLEIEL 753
QY 683 KNEFKFKKVTYPCAGKPLPHIIGSDLIHAKHAKNTELEEMLRQEMEVQONHAKKEES 738
Db 754 SNLK-----AELLS--VKQOLETE---REEKELKREAKENT 785
RESULT 8
US-08-923-992A-2
; Sequence 2, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-Iga Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996

```
ATTORNEY/AGENT INFORMATION:
: NAME: Esmond, Robert W.
: REGISTRATION NUMBER: 32,893
: REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1164 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: us-08-923-992A-2

Query Match          4.08; Score 157; DB 4; Length 1164;
Best Local Similarity 20.19; Pred. No. 0.00049;
Matches 144; Conservative 121; Mismatches 236; Indels 216; Gaps 40;

Qy 192 QIESFYPLDPSIGSKNVDLSGR-----QERKQIFKGTFFLNAKQHKLSSA 241
Db 131 QKNEFKTIDETNSDALLELENFNETNRLHLIKQHEEVEKDK-----KAKQOKTLKQS 185
Qy 242 ---VFGGGEARLITEENEEHNFPLAPGTCVDTGINTSOTLIPDCOKKWIQSIMDLQ 298
Db 186 DTKVDLSNIDKELNHQKSOVE-----KMAEQKGITNED---KDSMLKKIEDIRKQAO 234
Qy 299 ROGRLRPIEAEI--GLAVIEMTTKNYCDP--QGHPSGCTKTTTPGPSLSOGVSVDEKLMPS 354
Db 235 QADKEDAEVREELKLFSTKAGLDQEOIEH-----VKETSSENTQ--KVDEHY-- 286
Qy 355 SAPVNTTYYVADTSEQADTWLSERPKEIKVSKME--QK-----FRMLS 399
Db 287 ---ANSQNLAKQSLEELDKATTNEQATQVKNOFLENAQKLEIQPLIKETNVKLYKAMS 343
Qy 398 QDATVRESCKTSSNNN--SMVSNLTAKMRIPNYQLSPYK-LPSINKSKDRASQ--QQQT 452
Db 344 ESLEQVEKELKHNEANLEDLVAKSKIEVREYEGKLNQSKNLPKELQLEEAHSLKQOV 403
Qy 453 NSIRNYF----QPSTKKR-ERD-----BENOEMSSCKSARIETSCSLLEQTOPATPPLWKN 503
Db 404 EDFRKKFKTSEQVTPKKRVKRDLAANENQO-----KIELTVS-----PENITVYEG 450
Qy 504 KEQHLSENEPVDTSNNLFTD--TDLKSIYKNSASKSHAABKLRNKKREMDVAIEDEV 562
Db 451 EDVAKFT----VTAKSDSKTTLDFSDL--LTKYNPSVS---DRISTNVKNTDNHKAIEIT 501
Qy 563 LEQL-----FKDTK---PELEIDVKVQKEEDVNVKRRPRMDIETNDFSDFAV 608
Db 502 IKNLKLNESQVTLKAKDSDGNVVEKFTTITVQKKEK-QVPKTP-----EQKDSKTEKV 556
Qy 609 PESSKISQENI-----GKKRELK-----EDSLW-SAKE--- 636
Db 557 PQEPKNDKNQLOELIKSAQOELEKLEKAIKELMEQPEIPSNPEYGIQKSIWESQKEPIQ 616
Qy 637 --ISNNDKLQDSS-----EMLPKLLL-----TEFR 660
Db 617 EAITSFKKIIGDSSSKYYTEHYFNKYKSDFMNYQLHAQMEMLTRKVQVYMNKYPDNAEIK 676
Qy 661 SLVIKNSTSRNPISNDYDGLKN-----FKKP-----KKVTPYAGK 698
Db 677 K-IFESDMKRTK---EDNYGLENDALKGFEYKFLTPFNKIKQIVDDLKKKVEDQDQAP 732
Qy 699 LPHIIGGSLTAHARKNTELEEWLROEMEVQNOH--AKEBSLADDLFRYPNPKRR 753
Db 733 IPE---NSEMDQAEKAKIAVSKVSKVLDGVHQLQKKNKSIYDLFLEKELEIKQO 786
```

RESULT 9

us-08-923-992A-10

: Sequence 10, Application US/08923992A

: Patent No. 6280738

```
GENERAL INFORMATION:
: APPLICANT: Tai, Joseph Y.
: APPLICANT: Blake, Milan S.
: TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
: TITLE OF INVENTION: Streptococcal Beta Antigens
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
: STREET: 1100 New York Avenue, N.W., Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/923,992A
: FILING DATE: 05-SEP-1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/024,707
: FILING DATE: 06-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Esmond, Robert W.
: REGISTRATION NUMBER: 32,893
: REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1164 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: us-08-923-992A-10
```

Query Match 4.08; Score 156; DB 4; Length 1164;

Best Local Similarity 19.48; Pred. No. 0.00059;

Matches 139; Conservative 125; Mismatches 239; Indels 212; Gaps 37;

```
Qy 192 QIESFYPLDPSIGSKNVDLSGR-----QERKQIFKGTFFLNAKQHKLSSA 241
Db 131 QKNEFKTIDETNSDALLELENFNETNRLHLIKQHEEVEKDK-----KAKQOKTLKQS 185
Qy 242 ---VFGGGEARLITEENEEHNFPLAPGTCVDTGINTSOTLIPDCOKKWIQSIMDLQ 298
Db 186 DTKVDLSNIDKELNHQKSOVE-----TMAEQLGITNED---KDSMLKKIEDIRKQAO 234
Qy 299 ROGRLRPIEAEI--GLAVIEMTTKNYCDPQGHPSGCTKTTTPGPSLSOGVSVDEKLMPS 356
Db 235 QADKEDAEVREELKLFSTKAGLDQEOIEHVKETT-----SEENTQKVDEHY----- 286
Qy 357 PVNTTYYVADTSEQADTWLSERPKEIKVSKME--QK-----FRMLSQD 399
Db 287 -PNSIQNLAKQSLEELDKATTNEQATQVKNOFLENAQKLEIQPLIKETNVKLYKAMS 345
Qy 400 APTVRESCKTSSNNN--SMVSNLTAKMRIPNYQLSPYK-LPSINKSKDRASQ--QQQTNS 454
Db 346 LEQVEKQLKHNSQANLEDLVAKSKIEVREYEGKLNQSKNLPKELQLEEAHSLKQOVVED 405
Qy 455 IRNYF---QPSTKKR-ERD---EENOEMSSCKSARIETSCSLLEQTOPATPPLWKNKE 505
Db 406 FRKKFKTSEQVTPKKRVKRDLAANENQO-----KIELTVS-----PENITVYEGED 452
Qy 506 QHLSENEPVDTSNNLFTD--TDLKSIYKNSASKSHAABKLRNKKREMDVAIEDEVIE 564
Db 453 LKFT---LTAKSDSKTTLDFSDL--LTKYNPSVS---DRISTNVKNTDNHKAIEITIK 503
```

| | | | |
|---|-----|---|-----|
| Qy | 565 | QLFKDTRKPELEIDVK-----VQKQEDVNVYRKPRMDIETNDFSEAVPE | 610 |
| | | : : | |
| Db | 504 | NLKNESQVTLAAKDDSGNVVQKTFITIVQKKEK-QVPKTP-----EOKDSKTEEKVPQ | 558 |
| | | | |
| Qy | 611 | SSKISOENEIGK-----KRELKE-----DSLW-SAKE----- | 636 |
| | | : : : : : : | |
| Db | 559 | EPKSNKNOQLIELIKSAQOOLEKLEKAIKELMEQPEIPSNPEYGTQKSWESQKEPIQEA | 618 |
| Qy | 637 | ISNNDXLQDSS-----EMLPKKLL-----TEFRSL | 662 |
| | | : : : : | |
| Db | 619 | ITSFKKIIGDSSSKYYTEHFVNKYKSDFMNYQLHAQMEMLTRKVQYINKYPDNAEIKK- | 677 |
| | | : : | |
| Qy | 663 | VINKSTRNPNGINDYQOLKN-----FKKF-----KKVTPGAGKLIP | 700 |
| | | : : | |
| Db | 678 | IFESDMKRTK--EDNYGSLNDALKGYFEKFTLTPFNKIKQIVDDLOKKVQEQDQAPIP | 734 |
| Qy | 701 | HIIGSGDLIAHHARKNTELEEMLRQEMEVQNOH--AKEESLADDLFRYNPYLKR | 753 |
| | | : : : : : : : : | |
| Db | 735 | E--NSEMDOAKEKAIKSVKMSKVLIDGVGHLOKKNNNTKIVDLFKELEAIKQ | 786 |
| | | : : : : : : : : | |
| RESULT 10 | | | |
| US-09-214-564A-2 | | | |
| ; Sequence 2, Application US/09214564A | | | |
| ; Patent No. 6150515 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Sharp, Phillip A. | | | |
| ; APPLICANT: Zhou, Qiang | | | |
| ; TITLE OF INVENTION: TAT-SF: Cofactor For Stimulation Of Transcriptional | | | |
| ; TITLE OF INVENTION: Elongation By HIV-1 TAT | | | |
| ; FILE REFERENCE: M0656/7042 | | | |
| ; CURRENT APPLICATION NUMBER: US/09/214,564A | | | |
| ; CURRENT FILING DATE: 1999-08-18 | | | |
| ; PRIOR APPLICATION NUMBER: US 60/021,218 | | | |
| ; PRIOR FILING DATE: 1996-07-03 | | | |
| ; PRIOR APPLICATION NUMBER: US 60/033,152 | | | |
| ; PRIOR FILING DATE: 1996-12-13 | | | |
| ; PRIOR APPLICATION NUMBER: PCT/US97/11713 | | | |
| ; PRIOR FILING DATE: 1997-07-03 | | | |
| ; NUMBER OF SEQ IDS NOS: 5 | | | |
| ; SOFTWARE: FASTSEQ for Windows Version 3.0 | | | |
| ; SEQ ID NO 2 | | | |
| ; LENGTH: 754 | | | |
| ; TYPE: PRT | | | |
| ; ORGANISM: Homo sapiens | | | |
| US-09-214-564A-2 | | | |

```

RESULT 11
US-08-630-822A-68
; Sequence 68, Application US/08630822A
; Patent No. 5840695
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLENFELS, LYNDA
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
; TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,822A
; FILING DATE: 11-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CONNELL, GARY J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-17-C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 461 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-630-822A-68

```

Matches 93; Conservative 70; Mismatches 155; Indels 98; Gaps 21;

QY 270 VVDGTGITSOTLIPD-CQKKWIOQIMDLQROGLR-----PIPEAEIGLAVIFMTTKN 321
Db 43 MLDPLNPNQOLMFNYMQLQLOLQELHLSQQQPMHFEHPIPOE-----ATSTN 94

QY 322 YCDPQHPSTGLKTTTPGSPSLSQGVSVDEKMLPSAPV-NTTTYVADTSE-----QA 372
Db 95 Y-GPSGOYITSDATSYO-----STAOQFVFPQPIETTTTKIPETEIQIGVSNQYAO 145

QY 373 DTWDLSEPRKEIKVSKMEQKFR-----MLSDAPTVKESCKTSSNNNSMVSN 420
Db 146 ITNSNISPEVI-----GFRHYVAEQSGDVLHSHLHTEQPADKSTRGDQEPVSETG 198

QY 421 --LAKMRIPNYQLSPTK-LPSINKSKDRASQQOQTNSIRNYFOPSTKKRERDEENQEMSS 477
Db 199 SGFSYAQLSQGLKPTQPSNSVNLADRSPDLDTKTKENYKSP-----242

QY 478 CKSARIETCSLLEQTOPATPSL-WKNKEQHLSENEPVDNSNNLFTDLDKLSIVKNSA 536
Db 243 ---GRVQDITKIIDEKOKSKDTEWHNKK--VKEHKKVKD-----IKPDFESSORNNK 290

QY 537 SKSHAAEKLRSNKKREMDVAIEDEVL--EOLFKDTPKELEIDVKVQOEEDVNVKRPR 594
Db 291 SKN--IPKQIENTPQLDSLRSDIVIKGELLTKDTKSL--TTVNVDSLEDSVVKPKDEKP 347

QY 595 MDIETNDTFSDEAVPES-SKISQENEIGKKRELK-----EDLSWSA-KEISNNDK 642
Db 348 EPSEPSKTFDTSVAKDQVDNSTQANHHKKSKSPRKTETPEDEIEKALKETQASEK 403

RESULT 12

US-09-005-069-68
; Sequence 68, Application US/09005069
; Patent No. 5932470
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLENFELS, LYNDIA
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
; TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/09/005,069
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/630,822
; FILING DATE: 11-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: CONNELL, GARY J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-17-C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 461 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

; MOLECULE TYPE: protein
US-09-005-069-68

Query Match 3.8%; Score 150; DB 2; Length 461;
Best Local Similarity 22.4%; Pred. No. 0.00046;
Matches 93; Conservative 70; Mismatches 155; Indels 98; Gaps 21;

QY 270 VVDGTGITSOTLIPD-CQKKWIOQIMDLQROGLR-----PIPEAEIGLAVIFMTTKN 321
Db 43 MLDPLNPNQOLMFNYMQLQLOLQELHLSQQQPMHFEHPIPOE-----ATSTN 94

QY 322 YCDPQHPSTGLKTTTPGSPSLSQGVSVDEKMLPSAPV-NTTTYVADTSE-----QA 372
Db 95 Y-GPSGOYITSDATSYO-----STAOQFVFPQPIETTTTKIPETEIQIGVSNQYAO 145

QY 373 DTWDLSEPRKEIKVSKMEQKFR-----MLSDAPTVKESCKTSSNNNSMVSN 420
Db 146 ITNSNISPEVI-----GFRHYVAEQSGDVLHSHLHTEQPADKSTRGDQEPVSETG 198

QY 421 --LAKMRIPNYQLSPTK-LPSINKSKDRASQQOQTNSIRNYFOPSTKKRERDEENQEMSS 477
Db 199 SGFSYAQLSQGLKPTQPSNSVNLADRSPDLDTKTKENYKSP-----242

QY 478 CKSARIETCSLLEQTOPATPSL-WKNKEQHLSENEPVDNSNNLFTDLDKLSIVKNSA 536
Db 243 ---GRVQDITKIIDEKOKSKDTEWHNKK--VKEHKKVKD-----IKPDFESSORNNK 290

QY 537 SKSHAAEKLRSNKKREMDVAIEDEVL--EOLFKDTPKELEIDVKVQOEEDVNVKRPR 594
Db 291 SKN--IPKQIENTPQLDSLRSDIVIKGELLTKDTKSL--TTVNVDSLEDSVVKPKDEKP 347

QY 595 MDIETNDTFSDEAVPES-SKISQENEIGKKRELK-----EDLSWSA-KEISNNDK 642
Db 348 EPSEPSKTFDTSVAKDQVDNSTQANHHKKSKSPRKTETPEDEIEKALKETQASEK 403

RESULT 13

PCT-US93-03077-1
; Sequence 1, Application PC/TUS9303077
; GENERAL INFORMATION:
; APPLICANT: Board of Regents, The University of Texas System
; APPLICANT: Gaynor, Richard B.
; APPLICANT: Wu, Foon Kin
; TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR
; TITLE OF INVENTION: REGULATING GENE EXPRESSION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: PCT/US93/03077
; FILING DATE: 19930331
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,025
; FILING DATE: April 2, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Kammerer, Patricia A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: UTFD270PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1540
; TELEFAX: 713-749-2679

[illegible]

RESULT 14

```

US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DOEFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
;
US-08-353-700-1

Query Match 3.8% Score 146.5; DB 1; Length 3248;
Best Local Similarity 18.7%; Pred. No. 0.015;
Matches 152; Conservative 136; Mismatches 314; Indels 211; Gaps 33;

Qy 34 LIENDQSIQRNHAVLTANFSVTNLSQTDIEPVLTLKDNKSKYGTFFVNE-EKMQNGFSRTLK 92
| | | : : | | | : : | | | : : | | | : : | | | : : | | | : : |
Db 2129 LQEKLSQLEKDSQAL----SLTKCELENOIAQL----NKEKELLVKESESQARLSES-- 2178

Qy 93 SGDGIITGVGSGKRIEYEPVLVACSSLDVSGKLTALNOAILQLGFTYN-NKTEECTHLV 151
| | | : : | | | : : | | | : : | | | : : | | | : : | | | : : |
Db 2179 -----DYEK-----UNVS--RALEAALVEKGEFALRLSTQBEVHQL 2213

Qy 152 ---MVSVKVTIKTICALICGRPIVKPEYTFEFLKAVQSKQ--PPQIESFYPPDLDEPSIG 206
| | | : : | | | : : | | | : : | | | : : | | | : : | | | : : |
Db 2214 RRGIEKLRVIEA-----DEKQLHIAEKLRERENDSLKDKVENLERELQMSEEN 2265

Qy 207 SKNVDLSGRQERKQIFGKQKTFIFLNKQHKHKLSSAVVFGGEGEARLITENEEHNFFLAP 266
| | | : : | | | : : | | | : : | | | : : | | | : : | | | : : |
Db 2266 QELVILDAENSKAEVETLKTQIEEMARSLKIFELDLVTLRSEKENLTKIQIEKQ----- 2320

Qy 267 GTCVVDGTGITSOTLIPDCQKWKQ-----SINDMLQROGLRPIPEAEIGLA---VIFM 317
| | | : : | | | : : | | | : : | | | : : | | | : : | | | : : |
Db 2321 QLSELDLKLKSFLLSEKQAEIQIKESKTAVEMLNQ--LKELNVAALCGDQEIWK 2379

Qy 318 TTKNYCDPQGHPTGLKTTTPGPSLSQGVSVDEKL----- 352
| | | : : | | | : : | | | : : | | | : : | | | : : | | | : : |
Db 2380 ATFOSLDPIPEEHOIRNSI--EKI-RARLEFADEKKOICVLOOLKESHHADLLKGRVENI 2437

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:26:30 ; Search time 3146.82 Seconds
(without alignments)
18884.809 Million cell updates/sec

Title: US-09-837-602-1

Perfect score: 4403

Sequence: 1 ttccgacagggcggttg.....accggtggagctccagct 4403

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthba:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-------------|--------------------|
| 1 | 1460.2 | 33.2 | 1497 | 11 BC005293 | BC005293 Homo sapi |
| 2 | 1427 | 32.4 | 1461 | 11 BC016762 | BC016762 Homo sapi |
| 3 | 1391 | 31.6 | 2518 | 11 BC003719 | BC003719 Mus muscu |
| 4 | 1090.4 | 24.8 | 1792 | 11 BC013190 | BC013190 Mus muscu |
| 5 | 822.8 | 18.7 | 927 | 10 BG182890 | BG182890 RST1768 A |
| 6 | 741 | 16.8 | 789 | 10 BG214621 | BG214621 RST34264 |
| 7 | 732.4 | 16.6 | 902 | 10 BM461758 | BM461758 AGENCOURT |
| 8 | 725.4 | 16.5 | 775 | 10 BG676775 | BG676775 602623363 |
| 9 | 706.6 | 16.0 | 763 | 10 BM014420 | BM014420 603640169 |
| 10 | 696.2 | 15.8 | 960 | 10 BG284646 | BG284646 602408824 |
| 11 | 690.8 | 15.7 | 779 | 10 BG388866 | BG388866 602414591 |
| 12 | 687.8 | 15.6 | 712 | 9 AI796269 | AI796269 wh44g10.x |
| 13 | 676.2 | 15.4 | 735 | 10 BG779728 | BG779728 602668556 |
| 14 | 672.6 | 15.3 | 807 | 10 BG197194 | BG197194 RST16433 |
| 15 | 671 | 15.2 | 752 | 9 AU118357 | AU118357 AU118357 |
| 16 | 669.2 | 15.1 | 718 | 9 AW976050 | AW976050 EST388159 |
| 17 | 663 | 15.1 | 845 | 10 BG532783 | BG532783 602580405 |

| | | | | | | | |
|---|----|-------|------|------|----|----------|--------------------|
| c | 18 | 660.6 | 15.0 | 691 | 9 | AW978306 | AW978306 EST390415 |
| | 19 | 651.2 | 14.8 | 706 | 10 | BG292394 | BG292394 602386511 |
| | 20 | 648.4 | 14.7 | 690 | 10 | BG506358 | BG506358 601860423 |
| | 21 | 644 | 14.6 | 928 | 10 | BG392111 | BG392111 602410095 |
| | 22 | 633.4 | 14.4 | 686 | 10 | BG483074 | BG483074 602502895 |
| | 23 | 629 | 14.3 | 637 | 9 | AW237021 | AW237021 xm52b10.x |
| | 24 | 627.2 | 14.2 | 881 | 10 | BG109073 | BG109073 602281453 |
| | 25 | 621.2 | 14.1 | 667 | 10 | BF222791 | BF222791 7q23a10.x |
| | 26 | 612.6 | 13.9 | 642 | 10 | BG613414 | BG613414 602641314 |
| | 27 | 608.6 | 13.8 | 642 | 9 | AI888159 | AI888159 wm52h04.x |
| | 28 | 600.8 | 13.6 | 889 | 10 | BE786964 | BE786964 601476452 |
| | 29 | 596 | 13.5 | 604 | 9 | AW183153 | AW183153 x167b11.x |
| | 30 | 589 | 13.4 | 909 | 10 | BF027776 | BF027776 801764083 |
| | 31 | 587.4 | 13.3 | 1024 | 10 | BG202556 | BG202556 RST21915 |
| | 32 | 581.4 | 13.2 | 605 | 10 | BF511289 | BF511289 UI-H-B14- |
| | 33 | 576.4 | 13.1 | 672 | 10 | BG527527 | BG527527 602557140 |
| | 34 | 571.6 | 13.0 | 575 | 10 | BM312264 | BM312264 1q41c07.y |
| | 35 | 569 | 12.9 | 728 | 10 | BF382172 | BF382172 601814949 |
| | 36 | 564.2 | 12.8 | 651 | 9 | AV716658 | AV716658 AV716658 |
| | 37 | 559.2 | 12.7 | 872 | 10 | BF573240 | BF573240 602079623 |
| | 38 | 558.6 | 12.7 | 593 | 10 | BF219347 | BF219347 601884125 |
| | 39 | 554.8 | 12.6 | 573 | 10 | BE694454 | BE694454 OVO-BT070 |
| | 40 | 553.4 | 12.6 | 555 | 10 | BI962748 | BI962748 1e54b06.y |
| | 41 | 553.4 | 12.6 | 789 | 10 | BI691051 | BI691051 603309822 |
| | 42 | 543.8 | 12.4 | 636 | 9 | AW363125 | AW363125 CM2-CT031 |
| | 43 | 537.8 | 12.2 | 708 | 10 | BG530983 | BG530983 602560854 |
| | 44 | 537.6 | 12.2 | 757 | 10 | BF028917 | BF028917 601765293 |
| | 45 | 537.4 | 12.2 | 796 | 10 | BF248479 | BF248479 601821553 |

ALIGNMENTS

| | | | | | |
|------------|--|-------------|---|--------|-----------------|
| RESULT | 1 | 1497 bp | mRNA | linear | HTC 12-JUL-2001 |
| BC005293 | LOCUS | BC005293 | Homo sapiens, similar to nibrin, clone IMAGE:3997534, mRNA. | | |
| DEFINITION | BC005293 | BC005293 | HTC | | |
| ACCESSION | BC005293.1 | GI:14710042 | | | |
| VERSION | human, | | | | |
| KEYWORDS | Homo sapiens | | | | |
| SOURCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| REFERENCE | 1 (bases 1 to 1497) | | | | |
| AUTHORS | Strausberg, R. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (27-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA | | | | |
| REMARK | NIH-MGC Project URL: http://mgc.nci.nih.gov | | | | |
| COMMENT | Contact: MGC help desk Email: cgapbs-femail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: CLONTECH Laboratories, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www-shgc.stanford.edu Contact: (Dickson, Mark) mcdpaxil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M. | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1. .1497 | | | | |
| | /organism="Homo sapiens" | | | | |

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 16 Row: j Column: 6
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 3098674
This clone has the following problem: frame shifted.

```
/db_xref="taxon:9606"
/clone_image="3997534"
/tissue_type="Bladder, carcinoma"
/clone_lib="NIH_MGC_53"
/lab_host="DH10B"
/notes="vector: pDNR-LIB"
BASE COUNT      530 a      282 c      317 g      368 t
ORIGIN

Query Match      33.2%; Score 1460.2; DB 11; Length 1497;
Best Local Similarity 99.4%; Pred. No. 9.6e-212;
Matches 1476; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 22 acgtcgcccccagccctgagagcgccgaccgatgtgaaactgctgcccgcgcggggccc 81
DB 2  ACGTCGGCCCCAGCCCTGAGGAGCGGACCGATGTGGAACCTGCTGCCCGCGCGGCC 61

QY 82 ggcaggaggagaaccatacacagaacttttgactggcgttgagtacgtttgttggaaagaaaaa 141
DB 62  GCGAGGAGGAGAACCATACAGACTTTTGACTGGCGTTTGAGTACGTTGTGGGAAGGAAAA 121

QY 142 ctgtgcattcttaattgaaaatgatcagtcgatcagccgaatacatgctgtgttaactgc 201
DB 122 CTGTGCCATCTTAATTGNAATGATCAGTCGATCAGCGGAATCATGCTGTGTTAACTGC 181

QY 202 taacttttctgtaaccacactgagtaaacacagatgaaatccctgtattgacattaaaaa 261
DB 182 TAACCTTTCTGTAACCAACCTGAGTCAACACAGATGAAATCCCTGTATTGACATTAAAA 241

QY 262 taattctaagtatggtacccttggtaagtgaagaaaaaatgagaaatggcctttcccgaa 321
DB 242 TAATTCTAAAGTATGGTACCTTTGTTAATAGGAAAAAATGAGAAATGCTTTTCCCGAAC 301

QY 322 ttgaagtcggggatggtattacttttgagtggttggagtaaatcagaatagagta 381
DB 302 TTTGAAGTCGGGGATGGTATTACTTTTGGAGCTGTTTGGAAAGTAAATCAGAATAGAGTA 361

QY 382 tgagcctttggtgcagtcgtctcttctgttagatgctctgggaaaaactgctttaaata 441
DB 362 TGAGCCTTTGGTTCATGCTCTTCTGTTAGATGCTCTCGGAAAACTGCTTTAAATCA 421

QY 442 agctatatgcaacttgaggatttactgttaacaaatggacagaagaatgacctaccc 501
DB 422 AGCTATATTCGAACCTGGAGGATTTACTGTAAACAATTTGGACAGAAATGCACATCACT 481

QY 502 tgtcatggtatcagtgaaagtaccattaaaacaatatgtgcactcatttggacgtcc 561
DB 482 TGTCTATGATCAGTGAAGTTACCATTAACAACAAATATGTGCATCATTTGTGGCGTCC 541

QY 562 aattgtaaacgcagaatatttactgaattcctgaaagcagttcagttccaagaagcagcc 621
DB 542 AATTGTAAGCCAGCAATATTTACTGAATTCCTGAAAGCAGTTTCAGTCCAAAGAGCAGCC 601

QY 622 tccacaattgaaagtgttttaccacactcttgatgaaccatctatttggaaagtaaaatgt 681
DB 602 TCCACAATTTGAAAGTTTTTACCACACTCTTTGATGAACCATCTATTGGAAGTAAAAATGT 661

QY 682 tgatctgcagcgagcaggaagaaacaaatcttcaagggaacacatttatatttt 741
DB 662 TGATCTGTACGACGCGGAGGAAAGAAACAAATCTTCAAGGGAAACATTTATATTTT 721

QY 742 gaatgcaaacacagcataagaataggctccgcagttgtctttggaggtggggaagctag 801
DB 722 GAATGCCAAACAGCATAGAATTTGAGTTCGCGAGTTGCTCTTTGGAGTGGGGAAGCTAG 781

QY 802 gttgataacagaagagaatgaagaagaacataaatttttttggctccgggaacgtgtgt 861
DB 782 GTTGATACAGAGAGAATGAAGAAGAACATAATTTCTTTTGGCTCCGGGAACGTGTGT 841

QY 862 tgtgtatcaggagaatacaactcacagaccttaattcctgactgtcagaagaatgat 921
DB 842 TGTTGATACAGGAATAACAACCTCACAGACCTTAATTCCTGACTGTGACAGAAGAAATGGAT 901
```

```
QY 922 tcagtcaataatggtatgtctccaaagcaaggctcttagacctattctctgaagcagaat 981
DB 902 TCAGTCAATAATGGATATGCTCCAAAGGCAAGGCTTAGACCTATTCTCTGAAGCAGAAAT 961

QY 982 tggattggcgtgatttttcactgaactacaagaataactgtatctctcaggccatccag 1041
DB 962 TGGATTGGCGGTGATTTTCATGACTACAAAGAAATTTACTGTGATCTCTCAGGGCCATCCAG 1021

QY 1042 tacaggattaaagacaacaactccaggaccagcctttccaaagcgtgtcagttgatga 1101
DB 1022 TACAGGATTAAAGACAACAACCTCCAGGACCAGCCTTTCAAGGCGTGTCTAGTTGATGA 1081

QY 1102 aaaactaatgccaagcgcccgagtgaaactacaactacatcagtagtcacagaatcaga 1161
DB 1082 AAAACTAATGCAAGCGCCCGCAGTGAACACTACAACATACGTAGCTGACACAGAATCAGA 1141

QY 1162 gcaagcagatacatgggatttggattgaaagccaaagccaaagctctcaaaatga 1221
DB 1142 GCAAGCAGATACATGGGATTTGAGTGAAGGCAAAAGAAATCAAAAGTCTCCAAATGGA 1201

QY 1222 acaaaaattcagaatgctttcacaaagcgcacccactgtaaaggagctcgtcaaaacaag 1281
DB 1202 ACAAAATTCAAGATGCTTTTCAAGAGCAGCACCTGTAAGGAGTCTCTGCAAAACAAG 1261

QY 1282 ctctaaataatagtagtgatcaataactttggcttaagatgagaatcccaaaactatca 1341
DB 1262 CTCCTAATAATAATAGTAGTGGTATCAAAATACCTTTGGCTAAGATGAGAATCCCAAACTATCA 1321

QY 1342 gctttccaaactaaaattgccaagtataataaaaagtaaaagtagggctttctcagcaga 1401
DB 1322 GCTTTCCAACTAAATTTGCCAAGTATAAATAAAAGTAAAGTAGGGCTTTCTCAGCAGCA 1381

QY 1402 gcagacaaactcccatcagaaactacttccagcgtctacc-aaaaaagggaaggggatg 1460
DB 1382 GCACACCAACTCCATCAGAAACTACTTTTCAGCGCTCTACCAAAAGGGAAGGAGGATG 1441

QY 1461 aagaaatcagaagaatgtcttctcatgcaaatcagcagaagaatagaaa 1505
DB 1442 AAGAAATCAAGAAATGCTTCATGTCACAAAAAAGGAAAAAAGGAAAAA 1486
```

RESULT 2

```
BC016762      1460 bp      mRNA      linear      HTC 09-NOV-2001
LOCUS          Homo sapiens, Similar to Nijmegen breakage syndrome 1 (nibrin),
DEFINITION     clone IMAGE:4104186, mRNA.
ACCESSION      BC016762
VERSION        BC016762.1 GI:16876977
KEYWORDS       HTC.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1460)
AUTHORS        Strausberg, R.
TITLE          Direct Submission
JOURNAL        Submitted (05-NOV-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK         NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT        Contact: MGC help desk
Email: gcaps-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
```


[illegible]


```
Qy 577 atatttactgaattctctgaaagcagttcagttccaaagaagcagctccacaaattgaaag 636
Db 632 ATACTTTTCTGAATTTCTCAAGAGCAGTTGAATCTTAAGAAACAGCCTCCAGATATTGAAG 691
Qy 637 tttttaccacctcttaagaaacatatttggaagttaaaatattgtatctgtcaggacg 696
Db 692 TTTTATCCACCACCAATGATGAACAGCAGTATTGGAAGTAAAGTGTGATCTGTCCAGGCG 751
Qy 697 gcaggaagaagaaacaaattctcaagggaaacatttatatttttgatgccaacagca 756
Db 752 ACATGAAGAAGAAACAGATCTTCAAGGAGAAACATTTGTGTTTTAAATGCCAAGCAGCA 811
Qy 757 taagaattaggtcccgagttgtcttgaggagtggggaagctaggttgataacagaaga 816
Db 812 CAAGAAACTCAGCTCGCGAGTTGTTTCGGAGGTGGAGACCCAGGCTGATGGCAGAAGA 871
Qy 817 gaatgaagaagaacataattcttttggctccgggaacgtgtgttttgaatacagaat 876
Db 872 CGAGGAAGAGGAACAGAGCTTTCTTTCAGCTCCCGGAACCTTGCCTTGTGTGATGAGGAAT 931
Qy 877 aacaaactcacagaccttaattctctgactgtcagaagaatggattcagtcacataatgga 936
Db 932 AACGAATACACAGCTCATATTTTCACACTCCACAGAAATGGATTCATTTGATTAATGGA 991
Qy 937 tatgtccaaaggcaaggtcttagacctattctctgaagcagaatggattggcgtgat 996
Db 992 TACACTTCAAGGAATGCTTCAGACCTATTTCCTGAAGCGGAGATTTGGATTGGCTGTAT 1051
Qy 997 ttctactactacaagaattactgtatctctcagggccatcccgactcagagattaaagac 1056
Db 1052 TTTTATGACTACAGAGAAATTACTGTATCCGAGGGCCAGCCTTGTACAGAAATTAAAGAC 1111
Qy 1057 aacaactccaggaccagcctttcacaggcggtctcagttgataaaaaactaatgccaaag 1116
Db 1112 AACGACTCCAGGACCAGCCTTTCCCAAGTCTCTGTCGCCCAATGGAAAAATATCCCAAG 1171
Qy 1117 cgccccagtgaaactacaacatacgttagctgacacagaatcagagcaagcagatacatg 1176
Db 1172 TGCTCCAGTGAATATGACCATACATACGTAGCTGACACAGAAATCAAGAGCAGCAGATACATG 1231
Qy 1177 ---ggatttgagtgaaagggccaaagaaatacaagtgctccaaatggaaacaaatcag 1233
Db 1232 TATGCCCTTTGAGTGAAGACCAAGAAAGTAAAGATCCCTGGACTGGAACAAAGCTCTAG 1291
Qy 1234 aatgctttcacaagcagccaccactgttaaaggagtcctgcacaaacaaagcgtctataataa 1293
Db 1292 GAAACTTTTCAACAAGAAACATTCATATATAAGAGGAGGCCCTAAACCAAGCTCCAAAGCTAA 1351
Qy 1294 tagtatggtatcaataactttgggctaagatgagaatcccaactatcagctttcaccaac 1353
Db 1352 CAACGTAGCATCAGATACGCTGGTTAGAGGAAGAACCCCGAGCTATCAGCTTTCTCCAAT 1411
Qy 1354 taaattgccaagtataataaaagtaagaatagggtctctcagcagcagcagacacac 1413
Db 1412 GAAATTTCTCTTCCAAATAAATAAGGATTGGACTTCTCAGCAGCAGCAG---AACTC 1468
Qy 1414 catcagaactactttcagcgcgtctacacaaagggaaagggatgaagaaatcaaga 1473
Db 1469 CATCAAAAACCTACTTCCAGGCCATGCACCAGAAAAAGGAAAGGGATGAAGACAACCCAGA 1528
Qy 1474 aatgtcttcagcaaatcagcaagaatagaaacgtctgttctcttttagaacaacacaca 1533
Db 1529 GCTCTCTCTGTCGTAATTCATCCAGATGGAGCTCTCTCTGTCTCTTTTAGAACAAACACA 1588
Qy 1534 acctgtcacacctcattgttgaaaaataaaggagcagcatctatctgagaatgagcctgt 1593
Db 1589 ACCTGCCGGACCTTCACTGTGGAAAAAGCAAG---AGCATCAGTCTCAGAAATGGACCCCT 1645
Qy 1594 ggacacaaactcagacaataacttattacagatacacagatttaaaatctatttgaaaaa 1653
Db 1646 GGACAGGGAAGCCGACACCTCATCTGTGGTGGGATGGACATAGAACTCAACAGGAAGAG 1705
Qy 1654 ttctgccagtaaatctctatgctgcagaaaaagcctaagatcaataaaaaaaggggaaa 1709
```

```
Db 1706 TCCTGACAGGAACCACTTCCACAGAAACTCTTAGACCCAGAAAAAATAAAAAA 1761

RESULT 5
Bg182890
LOCUS Bg182890 927 bp mRNA linear EST 21-APR-2001
DEFINITION RST1768 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION Bg182890
VERSION Bg182890.1 GI:13704577
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 927)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
TITLE Creation of genome-wide protein expression libraries using random
activation of gene expression
J. Nat. Biotechnol. 19 (5), 440-445 (2001)
JOURNAL
MEDLINE 2127151
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave., Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scai@atersys.com
High quality sequence stop: 562.
FEATURES
Location/Qualifiers
1..927
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 330 a 190 c 187 g 214 t 6 others
ORIGIN

Query Match 18.7%; Score 822.8; DB 10; Length 927;
Best Local Similarity 94.7%; Pred. No. 3.3e-115;
Matches 859; Conservative 0; Mismatches 47; Indels 1; Gaps 1;
```

```

Db 321 TTTGGCTCCGGGACGCTGTTGTTGATACAGGAATAACAAACTACAGACCTTAATTC 380
QY 901 tgactgtcagaataatgagtcagtcataatgatatctcctcagaagcaggtcttag 960
Db 381 TGACTGTGCAAGAAATGGGATTCAGTCAATTAATGGATATGCTCCAAAGGCAAGCTCTAG 440
QY 961 acctattcctgaagcagaataatgattggcggtgatttctatgactacaaagaattactg 1020
Db 441 ACCTATTCTTGAAGCAGAAATGGATTGGGTGGGTGATTATCATGACTACAAAGATTAATG 500
QY 1021 tgatctcaggggccatccccagtcacaggattaaagacaaactccagaccagccttc 1080
Db 501 TGATCTCTCAGGGCCATCCCATGACAGGATTAAGACAACTCCAGGACCAAGCCCTTC 560
QY 1081 acaaggcgtgtcagttgatgaaaaactaatgccaagcgccccagtggaactacaacata 1140
Db 561 ACAAGGCGTGTGAGTGAAGAACTAATGCCAAGCGCCCATGAGTGAACACTACAACATA 620
QY 1141 cgtagctgacacagaatcagacgaagcagatatacatggtgattgagtaaaaggccaaaga 1200
Db 621 CGTAGCTGACACAGAATCAGACGAAGCAGATACATGGGATTTGAGTGAAGGCCCAAGA 680
QY 1201 aatcaaatgtccaaatggaacaaaaatcagaatgcttttcacaaagacacacccactgt 1260
Db 681 AATCAAGTCCCAAAATGGAACAAAAATTCAGAAATGCTTTTCAAGACGCACCCACTGT 740
QY 1261 aaaggagctctgcaaaacagctctaataataatagatagtcatactgtaactcttgctaa 1320
Db 741 NAAGGAGTCTGCANATAAGCTCTAAT-ATTAATAGTATGGGATCAATACTCTGGGTAA 799
QY 1321 gatgagaatccccaaactcagctttccaccaactaaattgccaaagtataataaagtaa 1380
Db 800 GATGAGAATNCCAACTACTCAGCTTTCCACCACTTAATTCGCCAAGTATTAATAAGTA 859
QY 1381 agatagggtcttcagcagcagcagacccaactccatcagaaactacttcagcogtctac 1440
Db 860 AGATAGGGCCTCTCAGCAGCAGAGAACCACTCCATCAGAAACCTTTCAGGCGCTTAC 919
QY 1441 caaaaa 1447
Db 920 CACAAA 926

RESULT 6
LOCUS BG214621 789 bp mRNA linear EST 21-APR-2001
DEFINITION RST34264 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG214621
VERSION BG214621.1 GI:13740642
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 789)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
E., Veloso,N., Klika,A., McElligott,K., Boozer,S., Mays,R., Smith
J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
JOURNAL 2127151
MEDLINE Contact: Scott J. Cain
COMMENT Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scai@atersys.com
High quality sequence stop: 545.
Location/Qualifiers
```

```

source 1..789
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 284 a 156 c 189 g 189 t
ORIGIN

Query Match 16.8%; Score 741; DB 10; Length 789;
Best Local Similarity 97.8%; Pred. No. 8.3e-103;
Matches 772; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

QY 637 tttttacccacctcttgatgaac-cattctattggaagtataaaatgttgatctgcaggac 695
Db 2 TTTTATTACCACCTCTTGTATGAACATCATCTATTGGAAGTAAAAATGTTGATCTGTCAAGAC 61
QY 696 ggcaggaaagaaacaaatcttcaaaagggaacattttatttttgaatgccaaacagc 755
Db 62 GGCAGGAAGAAAAACAATCTTCAAGGGAAAAACATTTATATTTTGAATGCCAAACAGC 121
QY 756 ataagaattgagttccgcagttgtctttggaggtgggaagctaggttgataacagaag 815
Db 122 ATAGAAATTTGAGTTCCTCCAGTGTCTCTTTGGAGGTGGGGAGGCTAGGTTGATATAACAGA 181
QY 816 agaataagaagaacataaattcttttgcctccgggaacgtgtgtgttgatcacagaa 875
Db 182 AGATGAAGAGAACAATAATTTCTTTTGGCTCCGGGAACGCTGTGTGTTGATACAGGA 241
QY 876 taacaaactcacagacetttaattcctgactgtcagaagaatggattcagtcataatgg 935
Db 242 TAACAAACTCACAGACC-TGATTTCTGACTGTCAGAGAAATGGATTTCAGTCAATAATGG 300
QY 936 atatgtccaaaggcaaggctcttagacctattcctgaagcagaattggattgctggtga 995
Db 301 ATATGCTCCAAAGCAAGGCTTTAGACCTATTCTCTGAAGCATATAATTTGGATTGGCGGTGA 360
QY 996 ttttcagtactacaaagaattactgtgctcctcagggtccatcccagtlacaggtataaga 1055
Db 361 TTTTCATGACTACAAAGAAATTTACTGTGATCCTCAGGGCCATCCAGTACAGGATTAAGA 420
QY 1056 caacaactccaggaccagcctttcacaaggcgtgtcagttgataaaactaatgccaa 1115
Db 421 CAACAACCTCCAGGACCAGGCGCTTTTCACAAGGCGTGTCTGATGATGAAAAACTAATGCCAA 480
QY 1116 gcgccccagtgaaacactacacatacgtagctgacacagaatcagacagacagatacat 1175
Db 481 GCGCCCCAGTGAACACTACAACTACATACGTAGCTGACACAGAAATCAGCCGACAGCATACAT 540
QY 1176 gggatttgagtgaaaggcccaaaagaatacaaaagtctccaaaatggacaaaattccagaa 1235
Db 541 GGGATTTGAGGGAAGGCCCAAAAGAAATCAAAAGTCTCCAAATGCGACAAAAAATTCAGAA 600
QY 1236 tgccttcacaagaagcagccaccctgtaaaggagtcctgtcgaacacagctcttaataata 1295
Db 601 TGCTTTTCAAGAGCGCACCCACCTCTTAAAGGAGTCTCTGCAAAACAAGCTCTAATAATAA 660
QY 1296 gtatggtatcaaatacttttggttaagatgagaatcccaactatcagctttccaccaacta 1355
Db 661 GCATGGTATCAAAATACCTTTGACTAAGATGAGAAATCCCATCTACTCATCTTTTCACCAACTA 720
QY 1356 aattgccaagtataataaagaataagtagggcttctcagcagacagacacacactcca 1415
Db 721 AATTGCCAAGTATAAATAAAGTAATGATAGGGCTTCTTCATCAGCAGACGACCAACTCCA 780
QY 1416 tcagaaact 1424
||||| |||
```

[illegible]

| Best Local Similarity | | 99.38; | Pred. No. 1.9e-100; | Matches 748; Conservative | | 1; | Mismatches | 2; | Indels | 2; | Gaps | 2; |
|-----------------------|---|---|---------------------|---------------------------|-----|-------------|------------|----|--------|----|------|----|
| Qy | 2637 | tactatttttkaataaatttctctacacaaatgtagcataacatcatgctcagtggttcttacac | 2696 | | | | | | | | | |
| Db | 24 | TTCTCATTTTTTAATATTTCTCTACACAAATGATAGCATACATATGCGAGTGTCTACAC | 83 | | | | | | | | | |
| Qy | 2697 | cttgcttttttacttagtaagattaaaaattataggaaatatcaataataatgtttttaata | 2756 | | | | | | | | | |
| Db | 84 | CTTGCTTTTTTACTTAGT - AGATTAAAAAATTATAGGAATATCAATATAATGTTTTTAATA | 142 | | | | | | | | | |
| Qy | 2757 | ttttttcttttccaattatgctgttagcttaactaaaaactctgggtgatccaaacaaatggc | 2816 | | | | | | | | | |
| Db | 143 | TTTTTCTCTTTTCCAAATATGCTGTAGCTTACCTAAACTCTGGTGATCCAAACAAATATGC | 202 | | | | | | | | | |
| Qy | 2817 | ttaagtggtcagatgtccacctacatgtattctctagttacttagaagaactggaacacatgtgg | 2876 | | | | | | | | | |
| Db | 203 | TTCAAGTGGTGCAGATGTCACCTACATGTTATCTTAGTACTAGAAGCTGAAGACCATGTGG | 262 | | | | | | | | | |
| Qy | 2877 | agacttcatcaaacatgggttttagttttccaccagaatggaaagacctgtacccctttttg | 2936 | | | | | | | | | |
| Db | 263 | AGACTTCATCAACATGGGTTTACTTTTCACCAAGATGGAAGACCTGTACCCCTTTTG | 322 | | | | | | | | | |
| Qy | 2937 | gtggtcttactgaactgggtgggtgtc-tgttttgaccttatattagagtcctagtatttcc | 2995 | | | | | | | | | |
| Db | 323 | GTGGTCTTTACTGAGCTGGGTGGGTCTGTGTTTGGAGCTTATTTAGAGTCTCTAGTTTCC | 382 | | | | | | | | | |
| Qy | 2996 | tacttataaagtgaagaatggtgaagatgtttctttcttcttctaccktaaaaggagatgtaaa | 3055 | | | | | | | | | |
| Db | 383 | TACTTATAAAGTAGAAATGGTGAGATTGTTTCTTTTCTACCTTAAAGGGAGATGGTAA | 442 | | | | | | | | | |
| Qy | 3056 | gaacaatgaatgtcttttttccaaacttatgtgacaagtgtatttcaagctgtgtgtcaa | 3115 | | | | | | | | | |
| Db | 443 | GAAACAATGAATGCTTTTTCACACITTTATTGACAGTGTATTTTCAAGTCTGTGTTCAA | 502 | | | | | | | | | |
| Qy | 3116 | aaatatattcatgttacctgtgtatccagcaagaaggagggttccagtccaagagtcactcaa | 3175 | | | | | | | | | |
| Db | 503 | AAATATATTTCATGTACCTGTGTGATCCAGCAGAAAGGGAGTTCACGTCAAGAGTCACTACA | 562 | | | | | | | | | |
| Qy | 3176 | ctgattagtgttttagaagaatgaagaaatggaacagtgaaggaatgaaggccatatattccat | 3235 | | | | | | | | | |
| Db | 563 | CTGATTAGTGTCTTTAGAGAATGGAATATGGAACAGTGAAGGAATGGAGGCCATATTTCCAT | 622 | | | | | | | | | |
| Qy | 3236 | gacttccctgttaaacagaagcaacagaaggaggaagaggtgctctcatcacctctc | 3295 | | | | | | | | | |
| Db | 623 | GACTTCCCTTGTAAACAGAACACACAGAGGACAGAGGCTGCGCTCTACATCACTCTC | 682 | | | | | | | | | |
| Qy | 3296 | accttccaaatcttgttgaagtgcattacttgcagaaacaaataaacttacttccaaag | 3355 | | | | | | | | | |
| Db | 683 | ACCTTCCAAATCTTGTGGAGTGTGATCTACTTGCAGAACCAAAATTAACTTACTTCCAAG | 742 | | | | | | | | | |
| Qy | 3356 | tcttgctgtcttcaggtgaacctccagctgca | 3388 | | | | | | | | | |
| Db | 743 | TTCTGGCTGCTTGGAGGTGGAACCTCCAGCTGCA | 775 | | | | | | | | | |
| RESULT | 9 | | | | | | | | | | | |
| LOCUS | BM014420 | | | | | | | | | | | |
| DEFINITION | 6036401069F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5416408 5', | 763 bp | mRNA | linear | EST | 30-OCT-2001 | | | | | | |
| ACCESSION | BM014420 | | | | | | | | | | | |
| VERSION | BM014420.1 | | | | | | | | | | | |
| KEYWORDS | EST. | | | | | | | | | | | |
| SOURCE | human. | | | | | | | | | | | |
| ORGANISM | Homo sapiens | | | | | | | | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | | | | | | | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | | | | | | | | |
| TITLE | 1 (bases 1 to 763) | | | | | | | | | | | |
| JOURNAL | NIH-MGC http://mgc.nci.nih.gov/ . | | | | | | | | | | | |

```
Db 600 AATGAATCGGGGAAGAAACGTGAACCTCAAGGAAGACTCACTATGGTCAAGTAAAGAAAT 659
Qy 1963 atctaaacatgacaaacttcaggatgatagatgagatgcttcc-aaaaaacgtgtattga 2021
|||||
Db 660 ATCTAACAACTGACAACTTCAGGATGATAGTGAATCTCCAAAAAAGCTGTCATCGA 719
Qy 2022 ctgaa-tttagatcactggtgattaaaaa 2049
|||||
Db 720 CTGAATCTTAGATCACTGGTGATTAAAAA 748

RESULT 10
Bg284646
LOCUS 602408824F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4537973 5',
DEFINITION mRNA sequence.
ACCESSION BG284646
VERSION BG284646.1 GI:13035811
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 960)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM10463 row: p column: 06
High quality sequence stop: 771.
Location/Qualifiers
1..960
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4537973"
/clone_lib="NIH_MGC_91"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 328 a 185 c 220 g 227 t
ORIGIN

Query Match 15.8%; Score 696.2; DB 10; Length 960;
Best Local Similarity 99.1%; Pred. No. 4.8e-96;
Matches 742; Conservative 0; Mismatches 3; Indels 4; Gaps 4;

Qy 517 gaaagttaccattaaaacaatatgtgactcatttggagcgtcccaattgtaagccaga 576
Db 1 GAAAGTTACCATTAAAACAATATGTGCACCTATTTGTGGAGCTCCAA-TGTAAAGCCAGA 59
Qy 577 atatttactgaattcctgaaacagttcagttccaagaagcagcctccacaattgaaag 636
Db 60 ATATTTTACTGAATTCCTGAAGCAGTTCAGTCCCAAGACAGCTCCACAAATTTGAAG 119
Qy 637 ttttaccacaccttggatgaaccattctattggaagtaaaaattgtgatctgcaggacg 696
Db 120 TTTTACCACCTCTTGTATGAACCATCTA-TGGAAGTAAAAATGTTGATCTGTCCAGGCG 178
Qy 697 gcaggagaagaaacaaattctcaagggaagaaacatttatatttttgatgccaacagca 756
|||||
```

```
Db 179 GCAGGAAGAAAAACAATCTTTCAAGGGGAAAAACAATTTATATTTTGAATGCCAACACGA 238
Qy 757 taagaaattgagttccgcagttgctcttggaggtggggaagctaggttgataacagaaga 816
|||||
Db 239 TAAGAAATTGAGTTCGCCAGTTGCTTTGGAGGTGGGAAGCTAGGTTGATAACAGAAGA 298
Qy 817 gaatgaagaagaacataaattcttttggctccgggaacgtgtgtgtttgtttgatacagaat 876
|||||
Db 299 GAATGAAGAAGAACATAAATTTCTTTTGGTCCGGGAACGCTGTGTTGTGTATACAGGAAT 358
Qy 877 acaaaactcacagacccttaattcctgactgtcagaagaataatggattcagttcaataatgga 936
|||||
Db 359 AACAAATCTCAGACACCTTAATTCCTGACTGTGAGAAGAAATGGATTGAGTCAATAATGGA 418
Qy 937 tatgtctcaaggaagctcttagacattctcctgaagcagaaattggattggcggtgat 996
|||||
Db 419 TATGCTCCAAAGGCAAGGCTTAGACCTATTCTTGAAGCAGAAATTTGATTGGCGGTGAT 478
Qy 997 ttctatgactacaaaagaattactgtgatctcctcaggggccatcccggtcacagattaaagac 1056
|||||
Db 479 TTTTCATGACTACAAAGAAATTAATCTGTGATCCTCAGGGCATCCCAAGTACAGGATTAAAGAC 538
Qy 1057 acaaaactccaggaccacccctttcacagcgctgtcagttgataaaaaactaatgccaaag 1116
|||||
Db 539 AACAACTCCAGGACCAAGCCTTTCCAAAGCGGTGTGATGATGAAAACTAATGCCAAG 598
Qy 1117 cgccccagtgaacactacaacatcagtagctgacacagaatcagacgaagcagatatacatg 1176
|||||
Db 599 CG-CCCAGTGAACACTACACATACGTAGCTGACACAGAATCAGACGAAGCAGATACATG 657
Qy 1177 gattttagtgaaaggccaaaagaatacaaaagtctcaaaatggaaacaaattcagaat 1236
Db 658 GGATTTGAGTGAAAGG-CAAAAGAAATCAAAAGTCTCCAACTTGGACAAACAAAATTCAGAAT 716
Qy 1237 gctttcaagaagcaccacctgtaaagg 1265
Db 717 GCTTTCAAGAGCCACCACCTGTAAGGG 745

RESULT 11
Bg388866
LOCUS 602414591F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4522774 5',
DEFINITION mRNA sequence.
ACCESSION BG388866.
VERSION BG388866.1 GI:13282312
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 779)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM10424 row: f column: 23
High quality sequence stop: 722.
Location/Qualifiers
1..779
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4522774"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
FEATURES
source
```


/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-qt primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
249 a 126 c 174 g 230 t

BASE COUNT
ORIGIN

Query Match 15.7%; Score 690.8; DB 10; Length 779;
Best Local Similarity 96.9%; Pred. No. 3.3e-95;
Matches 758; Conservative 0; Mismatches 17; Indels 7; Gaps 5;

QY 260 gataattcgaatgacgtacattgttaagagagaaatcagaaatgacgtttcccca 319
|||||
Db 1 GATAATCTAAGTATGGTACCTTTGTTAAGGAAATAATGCAATGCTTTCCCGA 60
|||||
QY 320 actttgaagtcg9ggagtgattacttttgagtggtttggaagtaaatcagaatagag 379
|||||
Db 61 ACTTTGAAGTCGGGGATGCTATTACTTTTGGAGTGTTCGAAGTAAATTCAGAATAGAG 120
|||||
QY 380 tatgagccttggttgatgctctcttggttagatgctctctgggaaactgctttaa 439
|||||
Db 121 TATGAGCCTTTGGTTGATGCTCTCTTGTGTTAGATGCTCTGGGAAACTGCTTTAAAT 180
|||||
QY 440 caagctatatgcaacttgagattactgttaacaaatgtggacagaaatgcactcac 499
|||||
Db 181 CAAGCTATATGCACTTGAGAGATTACTGTAACAAATTTGGACAGAAAGTGCATCAC 240
|||||
QY 500 ctgtgatggtatcagtgaaagtaccattataaaacaaatgtgcactcatttggagct 559
|||||
Db 241 CTGTGATGATCAGTGAAAGTTACCATTAAAAACAATATGTGCACATTTGTGGAGCT 300
|||||
QY 560 ccaatgtgaagccagaaatatttactgaattcctgaaagcagctcagtcgaagaagcag 619
|||||
Db 301 CCAATGTAAAGCCAGAAATATTTTACTGAATTCCTGAAAGCAGTTGAGTCAAGAACAG 360
|||||
QY 620 cctcccaaatgtgaagcttttaccacctcttgatgaaccatctattggaagtataatt 679
|||||
Db 361 CTCCCAAAATGNAAGTTTTTACCACCTCTTGTGATGAACCACTATTGGAAGTAAAT 420
|||||
QY 680 gttgatctgacgagcgaggaagaaacaaatcttcaaaagggaacacatttatatt 739
|||||
Db 421 GTTGATCTGTGAGCAGCGCAGGAAAGAAACAAATCTTCAAGGGAAACATTTATATT 480
|||||
QY 740 ttgaatgccaaacagcataagaattgagttccgcagttgtcttggaggtgggaaagct 799
|||||
Db 481 TTGAATGCCAAACAGCATAAGAAATTTGAGTTCCGCAGTTCTCTTTGGAGTGGGGAAGCT 540
|||||
QY 800 aggttgataacaga-agagaatgaagaagaacataattcttttggctccgggaacgtg 858
|||||
Db 541 AGGTTGTTACACACAGAGAAATGAAGNAGAACATAATTTCTTTTGGCTCCGGGAACGTG 600
|||||
QY 859 tgttggtgatacagggaataacaaactcacagacc-ttaattcctgactgtcagaagaat 917
Db 601 TGTGTTGATACAGGAATAACAACTCACAGACCTTTTAATTCCTGACTGTGAGAGAAAT 660
|||||
QY 918 ggattcagtcataatggaatgctccaaaggcaaggtc-ttagacctatctctgaagca 976
Db 661 GGATTACAGTCAAAATGGATATGCTCCAAAGGCAAGGCTCTTTAGACCTATTCTCGAAGCA 720
|||||
QY 977 gaaattggttggtggttttcatgactacaaagaattactgtgatctctcagggccat 1036
|||||
Db 721 GAA--TTGGATTGGGGGATTTTCATGACTACAA--GAATTCCTGGATCTCTCAGGGCAT 776
|||||
QY 1037 cc 1038
Db 777 CC 778

RESULT 12
AT1796269/c

LOCUS
DEFINITION
wh44q10.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2383650 3'
similar to TR:060672 060672 CELL CYCLE REGULATORY PROTEIN P95. [1]
// mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AT1796269
AT1796269.1 GI:5361732
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 712)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
REFERENCE
AUTHORS
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LENL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert length: 1177 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 451.
Location/Qualifiers
1. .712

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2383650"
/lab_host="NCI-CGAP_Kid11"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 154 a 127 c 148 g 281 t 2 others
ORIGIN

Query Match 15.6%; Score 687.8; DB 9; Length 712;
Best Local Similarity 98.0%; Pred. No. 9.6e-95;
Matches 695; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 987 tggcggtgattttcatgactacaaagaattactgtatcttcctcagggccatccccagtag 1046
|||||
Db 711 TTGCGGTGATTTTCTGACTACAAAGAATACTCTGTATCTNCAGGGCCATCNCAGTACAG 652
|||||
QY 1047 gattaagacaacaactccaggaccagcctttcacaggcgtgtcagttgatgaaaaac 1106
Db 651 GATTAAAGACAACAACATCCAGGACCAAGCCCTTTTCAAGGGCTGTCTGATTTGATGAAAAAC 592
|||||
QY 1107 taatgccaaagccccagtgaaactacaactacatagctgacacagaatccagacaag 1166
|||||
Db 591 TAATGCCAAGCGTCCAGTGAACACTAACATACGTAGCTGACACAGAAATCAGAGCAAG 532
|||||
QY 1167 cagatacatgggatttgagtgaaaggccaaagaaatcaaaagtctccaaatggaaacaaa 1226
|||||
Db 531 CAGATACATGGATCTGAGTGAAAGGCCAAAGAAATCAAAAGTCTCCAAATGGACAAA 472
|||||
QY 1227 aattcagaatgcttttcacagacgacccactgtaaggaggtcctgcaaaacagctcta 1286
|||||
Db 471 AATTGCAATGCTTTTCAAGACGACCCACTGTGTAAGGAGTCTCTGCAAAACAAAGCTCTA 412

| ORGANISM | REFERENCE | AUTHORS | TITLE | JOURNAL | MEDLINE | COMMENT |
|--|--|---|----------------|-------------|---------|---------|
| Homo sapiens | 513 | GTGATAGAGGCGATTAAGTTACAAATGTTTATGCGCTTANNATTATTAATAAATGCA | 572 | | | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | 2458 | caaaacttgatctcttctgtatgtaacaaatctgttctgtctgttccagcgttctgcatg | 2517 | | | |
| 1 (bases 1 to 807) | 573 | CAAACTTTGATGCTTTTGTGATGATGAACAATTCCTGTTCTGTTTTCAGGCTTTGTCATG | 632 | | | |
| Harrington, J. J., Sherf, B., Rundlett, S., Jackson, P. D., Perry, R., Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J., Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith, E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M. | 2518 | catcttttttcatctttttaaagtgtttgttttattataatagttatagtcacagttc | 2577 | | | |
| Creation of genome-wide protein expression libraries using random activation of gene expression | 633 | CATCTTTTTCATCTTNTTAAATGTTGTTTGGTTATTAAATAGTTAATATAGTCACAGTTC | 692 | | | |
| Nat. Biotechnol. 19 (5), 440-445 (2001) | 2578 | aaaactctaaatracgttaaggttaaggactaaagactaaccttcacacatgctcctagct | 2637 | | | |
| Contact: Scott J. Cain | 693 | AAAATTCATAATGTACGTAAAGGTAAA-GACTAAGTCAACCTTTCACATTTGCCGTAGCT | 751 | | | |
| Athersys, Inc. | 2638 | acttattt 2645 | | | | |
| 3201 Carnegie Ave, Cleveland, OH 44115, USA | 752 | ACTTGTT 759 | | | | |
| Tel: 216 431 9900 | | | | | | |
| Fax: 216 361 9596 | | | | | | |
| Email: scai@atersys.com | | | | | | |
| High quality sequence stop: 476. | | | | | | |
| Location/Qualifiers | | | | | | |
| 1. 807 | | | | | | |
| /organism="Homo sapiens" | | | | | | |
| /db_xref="taxon:9606" | | | | | | |
| /clone_lib="Athersys RAGE Library" | | | | | | |
| /cell_line="HT1080" | | | | | | |
| /note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances." | | | | | | |
| 291 a 130 c 140 g 239 t 7 others | | | | | | |
| BASE COUNT | | | | | | |
| ORIGIN | | | | | | |
| Query Match | 15.3% | Score 672.6; | DB 10; | Length 807; | | |
| Best Local Similarity | 96.6%; | Pred. No. 1.9e-92; | | | | |
| Matches 703; | Conservative | 2; | Mismatches 21; | Indels 2; | Gaps 2; | |
| QY 1919 | aaacgtgaactcaa-ggaagactcactatgttcagctcaaaataatctcaaatgacaa | 1977 | | | | |
| Db 33 | ACCCTGTAACCTCAAGGAAGACTCACTATGTCACCTAAGAAATATCTCAATGACAA | 92 | | | | |
| QY 1978 | acttcaggatgtagtgagatgcttcacaaaagctgttattgactgaatttagatcact | 2037 | | | | |
| Db 93 | ACTTCAGGATGATAGTGAGATGCTTCCAAAAAGCTGTTATGACTGAATTAGATCACT | 152 | | | | |
| QY 2038 | ggtgattaaaactctactctcagaaatcgtctgtgcatataatgatgattatggtcaact | 2097 | | | | |
| Db 153 | GGTGATTAATAACTCTACTTCCAGAAATCCATCTGGCATAATGATGATTATGTTCAACT | 212 | | | | |
| QY 2098 | aaaaattccaagaattcaaaaggtcacatatctctgagcaggaaacttcacacat | 2157 | | | | |
| Db 213 | AAAAATTTCAAGAAATTCAAAAGGTACATATCTCTGGAGCAGGAAACCTTCCACACAT | 272 | | | | |
| QY 2158 | cattgagagatcagatcctaatagtctcatctctcgaagaatacacagactagaagagt | 2217 | | | | |
| Db 273 | CATTGAGGATCAGATCTAATAGCTCATCATGCTCGAAGAATACAGAACTAGAGAGTG | 332 | | | | |
| QY 2218 | gctaaggcggaaatgagggtcaaaaaatcaaatgcaaaagaagtgctctctgtatga | 2277 | | | | |
| Db 333 | GCTAAGGCAGGAATGAGGATCAAAATCAACATGCAAAAGAAGAGTCTTTTGTGTATGA | 392 | | | | |
| QY 2278 | tcttttgaatacaatccttatttaaaaggagaagataaactgaggattttcaaaaaaag | 2337 | | | | |
| Db 393 | TCTTTTATAGATCAATCTCTATTTAAAAGGAGAGATAACTGAGGATTTTAAAAAAG | 452 | | | | |
| QY 2338 | ccatgaaaaacttctctagtagcactctacttcaggcccaacaaggttatatgaatatata | 2397 | | | | |
| Db 453 | CCATGGAAAACTTCTTAGTAGCACTACTTCAAGCCCAACAAAGTTATATGAATATATA | 512 | | | | |
| QY 2398 | gtgtatagaagcgtatttaagttagaattgttttatgtgcctaaattttattataaaaaatqca | 2457 | | | | |

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:27:55 ; Search time 5205.79 Seconds
(without alignments)
17699.435 Million cell updates/sec

Title: US-09-837-602-1

Perfect score: 4403

Sequence: 1 ttctgcacgagcgcggttg.....accggtgagctccagct 4403

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | ID | Description |
|------------|-------|-------|-------|--------|----|-------------|
|------------|-------|-------|-------|--------|----|-------------|

ALIGNMENTS

| RESULT | 1 | AF058696 | AF058696 | 4483 bp | mRNA | linear | PRI 20-MAY-1998 |
|------------|---|----------|----------|---------|------|--------|-----------------|
| LOCUS | | | | | | | |
| DEFINITION | | | | | | | |
| ACCESSION | | | | | | | |
| VERSION | | | | | | | |
| KEYWORDS | | | | | | | |
| SOURCE | | | | | | | |
| ORGANISM | | | | | | | |
| REFERENCE | | | | | | | |
| AUTHORS | | | | | | | |
| TITLE | | | | | | | |
| JOURNAL | | | | | | | |
| MEDLINE | | | | | | | |
| REFERENCE | | | | | | | |
| AUTHORS | | | | | | | |

Db 1561 ||||| TAAGGAGCAGCATCTCTGAGAAATGAGCCTGTGGACACAACTCAGACAAATAACTATT 1620
QY 1621 tacagatacagatttaaaatctattgtgaaatattccagtaaatctcattcctcagtcgaga 1680
Db 1621 TACAGATACAGATTTAAATCTATTGTGAAAAATCTCGCAGTAAATCTCATGCTCGAGA 1680
QY 1681 aaagctcaagatcaaaataaaaaagggaattggatgatgtggccatagaagatgaagattt 1740
Db 1681 AAAGCTAAAGATCAAAATAAAAAAGGGAATGGATGATGTGGCCATAGAAATCAAGTATT 1740
QY 1741 ggaacagttattcaaggacacaaaaaccagagttgagaattgtagtgaagttcaaaaaca 1800
Db 1741 GGAACAGTTATTCAAGGACACAAAAACAGAGTTAGAAATGTGCTGAAAGTTCAAAAACA 1800
QY 1801 ggaaggaagatgcaattgttagaaaaaggccaagatggatattggaatagaacaaatgacattt 1860
Db 1801 GGAGGAAGATGTCAATGTTAGAAAAAGGCCAAGGATGGATATAGAAACAAATGACACTTT 1860
QY 1861 cagtgatgaagcagtagccagaaagtagcaaaaatatctcaagaaaaatgaaattgggaagaa 1920
Db 1861 CAGTGATGAAGCAGTACCAGAAAGTAGCAAAATATCTCAAGAAAAATGAAATGGGAAGAA 1920
QY 1921 acgtgaactcaagggaagactcaactatggtcagctaaagaataatctaaacaatgacaaaact 1980
Db 1921 ACGTGAACCTCAAGGAAGACTCACTATGTGTCAGCTAAAGAAATATCTAACAATGACAACT 1980
QY 1981 tcagagatgagtgagatcttccaaaaaagctgttatgactgaatttagatcaactggt 2040
Db 1981 TCAGGATGATGATGAGATGCTTCCAAAAAGCTGTATTGACTGGAATTTAGATCACTGGT 2040
QY 2041 gattaaaaactctacttccagaataatcgtctggcataaaaatgattgattgtgcaactaaa 2100
Db 2041 GATTAAAACTCTACTTCCAGAAATCCGCTGCGCATAAATGATGATATGCTCAACTAAA 2100
QY 2101 aaatttcaagaaattcaaaaagctcacatatccttgagcagggaagaaacttccacacatcat 2160
Db 2101 AAATTTCAAGAAATTCAAAAGCTCACATATCCTTGAGCAGGAAAACTTCCACACATCAT 2160
QY 2161 tggagatcegatcctaagctcatcgtcgaagaatacagaactgagaagtggt 2220
Db 2161 TGGAGATCAGATCTAATAGCTCATCTGCTCGAAGAATACAGAACTAGAAGAGTGGCT 2220
QY 2221 aagcgggaaattgaggtcaaaatcaacatgcaaaagagctctctgtgctgatct 2280
Db 2221 AAGGCAGGAAATGGAGGTACAAAATCAACATGCAAAAGAAAGCTCTCTTCTGATGATCT 2280
QY 2281 ttttagatadaatccctatttaaaaagggaagaaatgagagattttaaagaagaagccca 2340
Db 2281 TTTTAGATACAATCCTTATTAAAAAGGAAGATAACTGAGGATTTTAAAAAGAGGCCA 2340
QY 2341 tggaaaaactcttagtagcatctacttcagcccaacagggttatgaatatagtg 2400
Db 2341 TGGAAAAACTTCTAGTAGCATCTACTTCAGGCCCAACAGGTTATGAAATATATAGTG 2400
QY 2401 tatagaagcagatttaagttacaattgtttatggcctaataatttataaataaaatgcacaa 2460
Db 2401 TATAGAAGCGATTAAAGTTACAATGTTTATGCGCTAAATTTATTAATAAATATGCACAA 2460
QY 2461 aacttggattcctttagtgaacaattgttctgtcgttttccagcgttctgcatgcat 2520
Db 2461 AACTTTGATTCTTTGATGTAACAATGTGTTGTTCTGTTTTCAGGCTTTGTCAATTGCAT 2520
QY 2521 cttttttcatttttaaaatgttttatttaaatagtttaatatagtcacagttcaaa 2580
Db 2521 CTTTTTTTCATTTTAAATGTGTTTTGTTTATTAAATAGTTAATATAGTACACAGTTCAAA 2580
QY 2581 attctaaatrtaaggttaaaggactaaagtcaaccttccaccattgtcctagctact 2640
Db 2581 ATTCTAAATGTACGTAAAGTAAA-GACTAAAGTCACCTTCCACCATTGTCCTAGCTACT 2639
QY 2641 ----- 2640

Db 2640 TGGTCCCTCCAGAAAAAAATTCATGACTACTCTTTCTTATGAATCTTTCCAGGATTTT 2699
QY 2641 -----Tattttaataattctcacaaatgatagcaataacata 2681
Db 2700 TGAGTCTCTATTCAAAATTCCTATTATTTTAAATTAATTTCTACAAATGATAGCATAA 2759
QY 2682 tgcagtggtctcacaccttgccttttacttagtaagaattaaaaattatagaatatcaat 2741
Db 2760 TGCAGTGTCTACACCTTGCTTTTACTTAGT-AGATTAAAAATATAGGAATATCAAT 2818
QY 2742 ataagtgttttaataatttttttccattatgctgtagcttaccataaactctggtga 2801
Db 2819 ATAATGTTTTTAATATATTTTCTTTTCCATTATGCTAGTCTTACCTAAACTCTGTGA 2878
QY 2802 tccaaacaaatggcttcagtggtgcagatgctcactcacatgttatctctagctagtaaaa 2861
Db 2879 TCCAAACAAATGGCTTCAGTGGTGCAGATGTCACCTACATGTTATTTCTAGTACTAGAA 2938
QY 2862 ctgaagaccatgtggagacttccataaacatgggttttagttttccaccagaatggaaagac 2921
Db 2939 CTGAAGACCATGTGGAGACTTCATCAAAACATGGGTTTAGTTTCCACCAGAAATGGAAGAC 2998
QY 2922 ctgtacccttttgggtggtcttactgagctgggtgggtgctctgttttggagcttattag 2981
Db 2999 CTGTACCCCTTTTGGTGGTCTTACTGAGCTGGGTGGGTGCTGTTTGGAGCTTATTTAG 3058
QY 2982 agtctagtttctactctataaagtagaaatgggtgagattgtttcttttctacckta 3041
Db 3059 AGTCTTAGTTTTCTACTATTATAAGTAGAAATGGTGAATGTTTTCTTTTCTTACCTTA 3118
QY 3042 aagggagatggtaagaacaatgaatgctcttttttccaaactttattgacaagtgttttc 3101
Db 3119 AAGGGAGATGGTAAGAAACAAATGAATGCTCTTTTTCAACTTTATTGACAAGTGATTTTC 3178
QY 3102 aagtcgtgttcaaaaatatattcaatgtaactgtgactgacccagcaagaaggaggttccagtc 3161
Db 3179 AAGTGTGTTCAAAAATATATTTCATGTACTGTGTATCCAGCAGAAAGAGGAGTTCAGTC 3238
QY 3162 aagagtcactcacactgattagttgttagagaatgagaatggaacagtggagaaatgga 3221
Db 3239 AAGAGTCACCTACAACTGATTTAGTTAGTGTAGAGAATGAGAAATGGAACAGTGGAGAA 3298
QY 3222 ggcataatttccatgacttcccttgttaacagaaagcaagaagggaagagctggcc 3281
Db 3299 GGCCATATTTCCATGACTTCCCTTGTAAACAGAAAGCAACAGAGGGAACAAGAGCTGGCC 3358
QY 3282 tctacatcaactcacaacttccaaactgttggaaagtgcactctacttccagaaacccaatt 3341
Db 3359 TCTACATCACTCTCACCTTCCAAATCTGTGGAAGTGCATCTACTTCCAGAACCAAAAT 3418
QY 3342 aacttacttccaaagtctgctgtcaggttggaactccagctgcaaggagttaggga 3401
Db 3419 AACTTACTTCCAGTCTGGCTGCTGTCAGGTGGAACCTCCAGCTGCAAGGAGTTAGGA 3478
QY 3402 aatgaagggtcttttttaaaagcttctcagccttccctagggaacagaaaaatgggtgagcc 3461
Db 3479 AATGAAGGTCTTTTAAAAAGCTTCTCAGCCTTCTTAGGGAACAGAAATTTGGGTGAGCC 3538
QY 3462 aactgtcaatttctactacaggcatgagaccagttagattattgaaatatattagagag 3521
Db 3539 AATCTGCAATTTCTACTACAGGCATGAGACCAGTTAGATTATTTGAAATATATTAGAGAG 3598
QY 3522 ttatgaacacttaaatattatgattgattgataattgataagaacatgggatactttaga 3581
Db 3599 TTATGAACACTTAAATTTATGATAGTGTATGACATTTGGATAGAACATGGGATACCTTTAGA 3658
QY 3582 agtagaattgacaggggcatattagttgataaaggagttcatttgagttctctttaaagcc 3641
Db 3659 AGTAGAATTGACAGGGCATATTAGTTGATGAATGGAGTCATTTGAGTCTCTTAATAGCC 3718
QY 3642 atgtataataattaccaagtggaagctgggtggaacatatggtctccattttacagtttaag 3701
Db 3719 ATGTATCATAAATTACCAAGTGAAGCTGGTGGACATATGGTCTCCATTTTACAGTTAAGG 3778

| | | | |
|------------|------|---|------|
| Qy | 3702 | aataatagacagattaataattgttctgtctatgcccacaatcccttttctaaggaaga | 3761 |
| Db | 3779 | AAATAATGACAGATTAATATTCTCTGTCTCATGCGCCACAATCCCTTCTTAAGGAAGA | 3838 |
| Qy | 3762 | ctgccctactacagctttttatatttgcgaatttgaataataatgaatgagagattc | 3821 |
| Db | 3839 | CTGCCCTACTATAGCAGTTTATATTGTCAATTTATGAATAATGAATGAGGAGTTC | 3898 |
| Qy | 3822 | tggctacctctgtctttacaaaattgggtgtgtccagctattttccctttttaaacmt | 3881 |
| Db | 3899 | TGGTACCTCCTGTCTTACAAAATATGGGTGTCTCCAGTATTTTCCCTTTTAACCAT | 3958 |
| Qy | 3882 | tcccaattcgggtgtgtaggtgagatgtttccatttgggttttaattgtatatccctgat | 3941 |
| Db | 3959 | T-CCAATCGGTGTGTAGTGATG- - - - -TTTCATTTTGGTTTAAATTGTATATCCCTGAT | 4012 |
| Qy | 3942 | agctataattgggtcatagaaattcttttatatacttctagatgaagctctctgtcggata | 4001 |
| Db | 4013 | AGCTATAATTGGGTGATAGAAAATCTTTATACATTTCTAGATGCAAGTCTCTTGTGCGATA | 4072 |
| Qy | 4002 | tacgtattgagatattacacacgtactgtgtgctgactgtttcttttatgtcttttgatg | 4061 |
| Db | 4073 | TATGTATTGAGATATTACACCTAGTCTGTGGCTTGACTGTTTCTTATGCTTTTGTATG | 4132 |
| Qy | 4062 | aatagaagttttaaatgttgacagaggttcaaatattattttttcttttggtttgatatattt | 4121 |
| Db | 4133 | AATAGAAGTTTAAATTTTGACAAGGTCAAAATTTATTTTCTTTTGTGTTGATATTTT | 4192 |
| Qy | 4122 | tctctcaatttaaccccaagatttcagatatctgtctctattatataaacctttatatatt | 4181 |
| Db | 4193 | TCTCTCAAAATTAACCCCAAGATTTCAGATATTCTGTCTTATTATATAAACTTTATATT | 4252 |
| Qy | 4182 | ttatattgtgactacgtgaattgatattgtattgttgtaattatgatacaggttctt | 4241 |
| Db | 4253 | TTATATTGTGATCTACTTGAATTGATATGATGTTGTGAATTATGATCAGGGTCTT | 4312 |
| Qy | 4242 | ttttcccccatacaagatccagctattgttaacactgttttattgaaagaattatccctt | 4301 |
| Db | 4313 | TTTTTCCCCCATACAAGTATCCAGCTATCTGTAACACTGTTTATTGAAAGAAATTATCTTT | 4372 |
| Qy | 4302 | cctcataaattacacttgccaatttagtaaaaaatcaataaacatrmammrrrgatc | 4361 |
| Db | 4373 | CCTCATTAATTAACCTTTGCCAAATTAGTCTGTCGCCGA- -ATTCCTGACGCCCGGGGATC | 4430 |
| Qy | 4362 | cactagtcttagagcgcgcaccgcggtgagctccagct | 4403 |
| Db | 4431 | CACTAGTCTTAGAGCGCGCCGCCCGGGTGGAGCTCCAGCT | 4472 |
| RESULT 2 | | | |
| AF051334 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| ACCESSION | | | |
| VERSION | | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| MEDLINE | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |

| | | | | | | | | | | |
|----------------------------|--|--|-----|--|--|--|--|--|--|--|
| JOURNAL | Submitted (26-FEB-1998) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr.11, Jena 07745, Germany | | | | | | | | | |
| FEATURES | Location/Qualifiers | | | | | | | | | |
| source | 1. .4412 | | | | | | | | | |
| | /organism="Homo sapiens" | | | | | | | | | |
| gene | /db_xref="taxon:9606" | | | | | | | | | |
| | 1. .4412 | | | | | | | | | |
| CDS | 52. .2316 | | | | | | | | | |
| | /gene="NBS" | | | | | | | | | |
| | /note="mutated in Nijmegen Breakage syndrome" | | | | | | | | | |
| | /codon_start=1 | | | | | | | | | |
| | /evidence=not_experimental | | | | | | | | | |
| | /product="nibrin" | | | | | | | | | |
| | /protein_id="AAC39732.1" | | | | | | | | | |
| | /db_xref="GI:3126795" | | | | | | | | | |
| | /translation="MKLLPAAGPAGGEPYRLLTGVVVGRKNCAILIENDQISIRN HAVLTANFSVTNLSQTDDEIPVFLKDNSKYGTFVNEEKMQNGFSRTLSKSGDITFGYF GSFRLEYELVACSCDLSVSGKALNQALILQUGFTVNNWTECTHLVMVSVKVTIK TICALTCGRDIVRPEYFEFLKAVESKKQPPQIESFYPLDPEPSIGKSNVDSGRQER KQIFKGTFTFLNAKHQKLSAVVFGGEGEARLITEENEHEHNFFLAGTCVDTGIT NSOTLIPDCOKWIOSIMDLORGLRPIPEAEIGLAVIEMTKNYCDPOGHPSTGLT TTTPGSLSGVSVDEKMLPSAPVNTTYYVADTSEQADTWLSEKPEIKVSKMEOK FRMLSQAPTVKESCKTSSNNNSMVSNTLAKMRIPNTQLSPKLPSPNSKSDRASQQQ QTSIRNYFQPTKKRREDEENQEMSSCKSARIETSCSLLEQTQTPATPSLWKNKEQHL SENEPVDTSNDNLFTDILKSIVKNSASKSHAAKRLSNKREMDDVAIIDEVLEQL FKDTKPELEIDVRVQOEEDVNVKRRPMIDETNDTFSDEAVPESISSQENETGKR ELKEDSLSAKESISNNDKLODDSEMLPKLLLTFRSLVTKNSTSRNPSGINDDYGOL KNFKPKKVTYPGAGKLPHTIGGSDLLAHARKNTELEEWLRQEMEQNHAKESLA DDLFRNPNYLRKR" | | | | | | | | | |
| polyA_signal | 2446. .2451 | | | | | | | | | |
| | /gene="NBS" | | | | | | | | | |
| | /evidence=experimental | | | | | | | | | |
| polyA_signal | 4392. .4397 | | | | | | | | | |
| | /gene="NBS" | | | | | | | | | |
| | /evidence=experimental | | | | | | | | | |
| BASE COUNT | 1470 a 753 c 853 g 1336 t | | | | | | | | | |
| ORIGIN | | | | | | | | | | |
| Query Match | 93.68; Score 4123.4; DB 9; Length 4412; | | | | | | | | | |
| Best Local Similarity | 97.18; Pred. No. 0; | | | | | | | | | |
| Matches 4287; Conservative | 7; Mismatches 28; Indels 91; Gaps 6; | | | | | | | | | |
| Qy | 13 | cgcggttgacgtcgcccgccagcctgagagccgacccgatgtggaactgctgccgc | 72 | | | | | | | |
| Db | 12 | CGCGGTTGACGTCGCCCCAGCCCTGAGGAGCCGCGGATGTGGAACCTGCTGCCGC | 71 | | | | | | | |
| Qy | 73 | cgcgggcccgcgagagagacacacagacttttgaactggcgctgagtaactgttgg | 132 | | | | | | | |
| Db | 72 | CGCGGCGCGGAGAGAGAACCATACAGACTTTTGTACTGGCGTTGAGTACGTTGTGG | 131 | | | | | | | |
| Qy | 133 | aaggaaaaactgtgccattcttaattgaaatgatcgatcgatcgacgcaaatcatgctgt | 192 | | | | | | | |
| Db | 132 | AAGGAAAACTGTGCCATTCTTGATTGAAATGATCAGTCGATCAGCCGAAATCATGCTGT | 191 | | | | | | | |
| Qy | 193 | gttaactgtctaaactttctgttaaccacactgagtcacacagatgaaatccctgtattgac | 252 | | | | | | | |
| Db | 192 | GTTAACTGCTAACTTTCTGTAAACCACTGAGTCAACACAGATGAATCCCTGTATTGAC | 251 | | | | | | | |
| Qy | 253 | attaaagataattctaaagtatgatactttgttaatgaggaataatgcagaatggctt | 312 | | | | | | | |
| Db | 252 | ATTAAGAGATAATCTAAGTATGTACTCTTTGTTAATGAGGAAAAATGCAGAAATGGCTT | 311 | | | | | | | |
| Qy | 313 | ttcccgaaactttgaagtcgggggagtgattacttttggagtgatttggaaataaattcag | 372 | | | | | | | |
| Db | 312 | TTCCCGAACTTTGAAGTCGGGGGATGTTACTTTTGGAGTGTGGGAAGTAATAATTGAC | 371 | | | | | | | |
| Qy | 373 | aatagatgatgcctttgttgatgctctcttctttagatgctctctgtggaaaaactgc | 432 | | | | | | | |
| Db | 372 | AATAGATGATGAGCCTTTGGTTGTCATGCTCTCTCTTTTGTAGATGCTCTCGGAAAACTGC | 431 | | | | | | | |
| Qy | 433 | tttaaatcaagctatattgcaacttggaggaatttactgtaaacaattggacagaagaatg | 492 | | | | | | | |

Db 432 TTTAAATCAAGCTATATTGCAACTTGGAGGATTTACTGTAAACAATGGACAGAGAATG 491
Qy 493 cactcacccttgcatggtatcagtgaaagttaaccattaaacaataatgtgcaactcaattg 552
Db 492 CACTCACCCTTGTCATGTTGATCAGTGAAGTACCATTAAACAATATGTGCACTCAATTTG 551
Qy 553 tggacgtccaattgtaaacgagcagaatattttactgaattccctgaaagcaggttcagtcsoa 612
Db 552 TGGACGTCCAATTGTAAGCCAGAGATATTTTACTGAATTCCTGGAAGCAGTTGAGTCCAA 611
Qy 613 gaagcagcctccaataattgaaagtttttaccacaacctcttgatgaaccatctattggaag 672
Db 612 GAAGCAGCCTCCACAAATTTGAAAGTTTTCACCACCTCTTGATGAACCATCTATTGGAA 671
Qy 673 taaaaatgttctgtcagcgagcgcaggaagaaacaaactctcaagggaacatt 732
Db 672 TAAAAATGTTGATCTGTCAGGACGGCAGGAAAGAAACAATCTTCAAAGGGAACAAT 731
Qy 733 tatattttgaatgccaaacagcataaagaatttgagttccgcagttgttcttggagtg 792
Db 732 TATATTTTGAATGCCAAACAGCATAGAATTTGAGTTCCGCAGTTGTCTTTGGAGGTGG 791
Qy 793 ggaactaggttgataacaagaagagaatgaagaacataatttcttttggctccgg 852
Db 792 GGAAGCTAGGTTGATAAACAGAAGAGAAATGAAGAACAATAATTTCTTTTGGCTCCGG 851
Qy 853 aacgtgtgtgttgatcacaggaataacaaactcacagaccttaattctcgtgctgcagaa 912
Db 852 AACGTGTGTTGTGATACAGGAATTAACAACATCACAGACCTTAATTTCTTGACTGTGAGAA 911
Qy 913 gaaatggattcagtcataatggatatgctccaaaggcagggtctcttagacctattcctga 972
Db 912 GAAATGGATTTCAGTCAATTAATGGATATGCTCCAAAGGCCAAGGTCATTAGACCTATTCCCTGA 971
Qy 973 aqcagaaattgattggtggtgatttctagctacaagaatactactgtgactcactcaggg 1032
Db 972 ACAGAAATTTGGATTGGCGGTGATTTCATGACTACAAAGAATTAAGTGTGATCCTCAGGG 1031
Qy 1033 ccattccagtcagatataagacaacaactccaggaccaagccttccagaaggcgtgctc 1092
Db 1032 CCATCCAGTACAGGATTAAGACACAACTCCAGGACCCAGCCCTTTCACAAAGGCGTGTC 1091
Qy 1093 agttgatgaaactaattgcgaagcccgccagtgaaactacaatacgtagctgacac 1152
Db 1092 AGTTGATGAAAACTAAATGCCAAGCCGCCAGTGAACACTACAACATACGTAGCTGACAC 1151
Qy 1153 aqaatcagaagcagagatacatgggatttgagtgaaagcccaaaagaataatcaaaagctc 1212
Db 1152 AGAATCAGAGCAAGCAGATACATGGGATTTGAGTGAAGGCCAAAAGAAATCAAAAGTCTC 1211
Qy 1213 caaaatggaacaaaaattcagaatgctttcaagaacgcaccactgtaaaggagtcctg 1272
Db 1212 CAAATGGAAACAAAATTCAGAAATGCTTTTCAAGAGTGCACCCACTGTAAAGGAGTCCGTG 1271
Qy 1273 caaaacagctctaataataatagtgatgatacaatactttggctaaagatgagaatccc 1332
Db 1272 CAAAACAAGCTCTAATAATAATAGTATGATATCAAAATCTTTGGCTTAAGATGAGAAATCCC 1331
Qy 1333 aaactatcagtttccacaactaaattggccaagataaaataaagtaagaatagggtctc 1392
Db 1332 AAACATATCAGCTTTCCACCAACTAAATTTGCCAAGTATAATAAAGTAAAGATAGGGCTTC 1391
Qy 1393 tcagcagcagcagacaactccatcagaactacttctcagcgctctaccaaaaaaggga 1452
Db 1392 TCAGCAGCAGCAGACCACTCCATCAGAAACATCTTTTCAGCCGCTTACCACAAAAGGGA 1451
Qy 1453 aaggatgaagaaatcaagaaatgcttctcatgcaaatcagcaagaatagaacgctcttg 1512
Db 1452 AAGGGATGAAGAAATCAGAAATGCTTCATGCAAAATCAGCAAGAAATAGAACGCTTTG 1511
Qy 1513 ttctcttttagaacaacacaacacctgctacacccctcattgtgaaaaataaaggagcaga 1572
|||||

Db 1512 TTCTCTTTTAGAAAAACACAAACCTGCTACACCCCTCATTTGTGGAAAAAATAAGGAGCAGCA 1571
Qy 1573 tctatctgaaatgacccctgtgacacaaactcagacaataaacttattttacacatcacaga 1632
Db 1572 TCTATCTGAAATGACCCCTGTGGACACAACACTCAGACAATAACTTATTTACAGATACAGA 1631
Qy 1633 tttaaatctattgtgaaaaattctctccagtaaatctctcgtcgcagaaaaagctaaagatc 1692
Db 1632 TTTAAATCTATTGTCAAAAATTTCTGCCAGTAAATCTCATGCTGCAGAAAAAGCTAAGATC 1691
Qy 1693 aaataaaaaagggaaatggatgagtggccatagaagatgaagtatttgggaaacagttatt 1752
Db 1692 AAATAAAAAAGGAAATGGATGATGTGCCATAGAAGATGAAGTATTGGAAACAGTTATT 1751
Qy 1753 caaggcacaaaccagagttagaatgatgtgaagttcaaaaaacagaggaagatgt 1812
Db 1752 CAAGGACACAACACAGAGTTAGAAATTTGATGTGAAGTTCAAAAACACAGGAGGAAGATGT 1811
Qy 1813 caatgttagaaaaagccaggatggatatagaaacaaatgcacacttccagtgatgaagc 1872
Db 1812 CAATGTTAGAAAAAGGCCAAGGATGGATATAGAAACAATAATGCACACTTTCAGTGATGAAGC 1871
Qy 1873 agtacacagaagttagcaaaatatctcaagaaaaatgaaattgggaagaaacgtgaactcaa 1932
Db 1872 ACTACCAGAAAGTAGCAAAATATCTCAAGAAATGAAATTTGGGAAGAAACGTGAACCTCAA 1931
Qy 1933 ggaagactcactatggtcagctaaagaaatatctcaaaatgacaaactgcacacttccagtgatgag 1992
Db 1932 GGAAGACTCATTATGTTGCTAGCTAAGAAATATCTACAAATGACAAACTTCAGGATGATAG 1991
Qy 1993 tgagatgcttccaaaaagcgtgttattgactgaatttagactcactcgtggtgattaaaaactc 2052
Db 1992 TGAGATGCTTCCAAAAAGCGTTATTGACTGAATTTAGATCACTGGTGATTAATAAACTC 2051
Qy 2053 tacttccagaataatccgtctggcataaataatgatgatattggtcacaactaaaaatttcaagaa 2112
Db 2052 TACTTCCAGAAATCCCATCTGGCATAAATGATGATTAATGTGTCACACTAAAAATTTTCAAGAA 2111
Qy 2113 attcaaaaagggtcacatatcctggagcaggaagaaactccacacatcatttgagagatcaga 2172
Db 2112 ATTCAAAAGGTCACATATCTCTGGAGCAGGAAACATTCACACATCATTTGAGGATCAGA 2171
Qy 2173 tctaaatagctcatcactgctgaaagaataacagaactagaaagatggtcgttaaggcaggaat 2232
Db 2172 TCTAATAGCTCATCATGCTCGAAGAAATACAGAACTAGAAAGATGGCTAAGCGCAGGAAT 2231
Qy 2233 ggaggtacaaaaatcaacatgcataaagaagatctctgtgctgatgactcttttttagatacaa 2292
Db 2232 GGAGGTACAAAATCAACATGCAAAAGAGAGTCTCTTGTCTGATGATCTTTTTTAGATACAA 2291
Qy 2293 tccttatttaaaaaggagaagataaactgaggaatttlaaaagaagccatggaacaaactc 2352
Db 2292 TCCTTATTTAAAAAGGAGAGATAACTGAGGATTTTAAAAAGAGCCATCGAAAAACTTC 2351
Qy 2353 ctagtgaagcatctacttcaggcccaacaaggttataagaatatagtgatagaagogat 2412
Db 2352 CTAGTAAGCATCTACTTTCAGGCCCAACAAGGTTATATGAATATATAGTGTATAGAAGCGAT 2411
Qy 2413 ttaagttacaatgtttttatggccctaaattttatttataataaataacacacaaactttgattct 2472
Db 2412 TTAAGTTACAATGTTTTTATGGCCCTAAATTTATTAATAAATGCACAAAACCTTTGATTTCT 2471
Qy 2473 ttgtgtgtaacaattgtttgtctgttttcaggctttgtcattgcatctcttttctcatt 2532
Db 2472 TTTGTATGTAACAATTTGTTTGTCTGTTTTCAGGCTTTGTTCATTGTCATCTTTTTTTCATT 2531
Qy 2533 tttaaatgtgtttgtttatttaaatagtttaatatagtcacagttcaaaaattcttaaaatrra 2592
Db 2532 TTTTAATGTGTTTGTTTTATTAATAATAGTTAAATATAGTCACAGTTCAAAATTTCTAAATGTA 2591
Qy 2593 cgtaaaggtaaggaactaaagtcacccctccaccattgtcctagctact----- 2640
Db 2592 CGTAAGGTAAA-GACTAAAAGTCACCCCTTCCACCATTGTCCTAGCTACTTGGTTCCTCCTCA 2650
|||||

```
Qy 2641 ----- 2640
Db 2651 GAAAAAATTCATGACTACTCAATTCATTATGAATCTTCCAGGGATTTTGTAGTCCATATPC 2710
Qy 2641 -----tattttfaataatttctcacacaaatgatagcataacatcagtggttcta 2693
Db 2711 AAATTCCTATATTTTAAATAAATTTCTACACAAATGATAGCATAACATATGAGGTGTTCTA 2770
Qy 2694 caccttgctttttacttagtaagataaataattataggaataatcaataatgttttta 2753
Db 2771 CACCTTGCTTTTTPACTTAGT - AGATTAAAAAATATAGGAATAATCAATATAATGTTTTA 2829
Qy 2754 atattttttcttccattatgctgtagcttacttaactaaactcctggtgatccaaacaaat 2813
Db 2830 ATATTTTTCCTTTTCCATTAATGCTGTAGTCTTACTTAAACTCTGGTGATCCAAACAAAT 2889
Qy 2814 ggtctcagtggtgcagatgtcacctcacatgcttattcttagtactagaacactgaagaccatg 2873
Db 2890 GGCTTCAGTGGTGAGATGTCACCTACATGTTATTCTAGTACTAGAAACGGAAGACCATG 2949
Qy 2874 tgagacattcatcaaacatgggttttagtttccacagaatggaaagacctgtaccccttt 2933
Db 2950 TGGAGACTTCATCAACATGGGTTTAGTTTTCACAGAAATGGAAGACCTGTACCCCTTT 3009
Qy 2934 ttggtgcttactagactgggtgggtgctgtgtttttagcttattagagtcctagtttt 2993
Db 3010 TTGGTGCTCTTACGTGAGCTGGGTGGGTGCTGTGTTTTCAGGCTTAATTTAGAGTCTCTAGTATTT 3069
Qy 2994 cctacttataaagttagaaatggtgagatgttttcttctttctaccktaaaaggagatggt 3053
Db 3070 CCTACTTATAAGTAGAAATGGTGAGATGTTTCTTTTCTACCTTAAAGGGAGATGGT 3129
Qy 3054 aagaacaatgaatgcttttttccaaacttattgacaagtatttccaagtctgtgttc 3113
Db 3130 AAGAAACAATGAATGTCTTTTTCAAACTTTATTGACAAGTGATTTTCAAGTCTGTGCTC 3189
Qy 3114 aaaaatatattcatgtacctgtcatccagcaagaaggagttccagttcaagagatcaactac 3173
Db 3190 AAAAATATATTCATGTACCTGTGATCCAGCAAGAAGGGAGTTCCAGTCAAGAGTCACTAC 3249
Qy 3174 aactgattagtgttagagaatgagaaatggaaacagtggaggaatggagcccatatttcc 3233
Db 3250 AACTGATTAAGTGTGTTAGAGAAATGAGAAATGGAACAGTGAAGGAATGAGGCCATATTTCC 3309
Qy 3234 atgaattcccttgaacagaagacacagaagggacaagaggtgacctctacatcactc 3293
Db 3310 ATGACTTCCCTTTCTAACAACAGAACCAACAGAAAGGCAAGAGGCTGGCCCTTACATCACTC 3369
Qy 3294 tcaccttccaaatcttgtggaagtgcactctacttgcagaaacaaataacttacttcca 3353
Db 3370 TCACCTTCCAAATCTTGTGGAAGTGATCTTACTTGCCAGAACCAAAATTAACCTTACTTCCA 3429
Qy 3354 agttctggtcgttcaggttgaactccagctcgaaggaggttagggaaatgaagtcctt 3413
Db 3430 AGTTCTGGCTGCTTGAGGTGGAACCTCCAGTCCAGGGAGTTAGGGAATGAGGTCCTT 3489
Qy 3414 tttttaaaagctctcagccttccctaggaagaacagaatgggtgagccaatctgcgaatt 3473
Db 3490 TTTTAAAAAGCTTCTCAGGCTTCTCAGGCAACAGAAATTTGGGTGAGCCAATCTGCAATTT 3549
Qy 3474 ctactacagccattgagaccagtttagattattgaaatattatagagagttatgaacactt 3533
Db 3550 CTACTACAGCATTTGAGACCAGTTAGATTATTGAAATATTTATAGAGAGTTATGAACACTT 3609
Qy 3534 aaattatgatgtggtatgacattggatagaacatgggatacttttagaagtgaattgac 3593
Db 3610 AAATTATGATAGTGGTATGACATTGGATAGAACATGGGATACCTTTTAGAGTAGAATTGAC 3669
Qy 3594 agggcatattagttgatgaaatggagtcattttagcttctttaaagccatgtatcataat 3653
Db 3670 AGGGCATATTAGTTGATGAAATGGAGTCACTTTGAGTCTCTTAAATAGCCATGTATCATAAT 3729
```

```
Qy 3654 taccaaatgaagctggttggaacatatggtctccatttttacagtttaagaataataatggac 3713
Db 3730 TACCAAGTGAAGCTGGTGGAAACATATGCTCTCCATTTTACAGTTTAAGGAATAATAATGGAC 3789
Qy 3714 agattaatatgttctgtctcatgccacaatccctttcttaaggaagactgccctactact 3773
Db 3790 AGATTAAATTTGTTCTCTGTCATGCCCAAAATCCCTTTCTTAGGAAGACTGCCCTTACTAT 3849
Qy 3774 agcagtttttattattgttcaatttatgaataataatgaatgagggagttctcgttacctcctg 3833
Db 3850 AGCAGTTTATATATTTGTCATTTATGAATTAATGAATGAATGA - GAGTTCTGGTACCTCCTG 3908
Qy 3834 tctttacaacatbattgggttggtccocagbatttttccctttttaacccmttcccaatccggg 3893
Db 3909 TCTTTACAAATATGCG-----TGTTGTCAGTATTTTTCCTTTTAAACCATTCCAATCGGTG 3964
Qy 3894 tgttagtggtgagatgttccatttgggtttttaaatttggatatccctcgatagctataattgg 3953
Db 3965 TGTAGTGATG-----TTTCATTTTGGTTTTAAATTTGTATATATCCCTGATAGCTATAAATGG 4019
Qy 3954 gtcatagaaaatcttttatcacattctagatgcaagctctcttgcggatatacgtattgaga 4013
Db 4020 GTCATAGAAATCTTTATACATTTCTAGATGCAAGTCTCTTCTCGGATATATGATGAGA 4079
Qy 4014 tattacacctagctgtggttgactgttttcttttatgtcttttttgatgaatagaagtttt 4073
Db 4080 TATTACACCTAGTCTGTGGCTTGACTGTTTCTTTATGTCCTTTTGATGAATGAAGAAGTTT 4139
Qy 4074 aaatttgaacaagtcataatttatttttcttttcttttggatttatttctcctcaattt 4133
Db 4140 AAATTTGACAAGGTCAAAATTTATTTTCTTTTGTGTGATATTTTCTCTCCAAATTT 4199
Qy 4134 aaccccaagatttcagatatattctctctattattataaaactttatttttatatttggga 4193
Db 4200 AACCCCAAGATTTCCAGATATTTCTGCTCTATTATATATAAATTTATATTTATATTTGGA 4259
Qy 4194 tctacctggaattgatatgtatgttgaattatggatcagacgggttcttttttcccccatt 4253
Db 4260 TCTACCTTGAATTGATATGTATGTGTGAATTTATGGATCAGGGTCTCTTTTTCCTCCCAT 4319
Qy 4254 acaagttacagtcattgtaacactgtttattgaaagaattatccttctcctcataaatt 4313
Db 4320 ACAAGTATCCAGTCATTTGTAACACTGTTTATTTGAAAGAAATATCTTCTCCCATTAATTT 4379
Qy 4314 accttgccaattagtaaaaaatcaatttaacct 4346
Db 4380 ACCTTGCCAATTAGTAAAAAAATCAATTAACCAT 4412
RESULT 3
AB013139
LOCUS
DEFINITION Homo sapiens gene for NBS1, complete cds.
ACCESSION AB013139
VERSION AB013139.1 GI:3169124
KEYWORDS NBS1; Nijmegen breakage syndrome.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (sites)
AUTHORS Matsura, S., Tauchi, H., Nakamura, A., Kondo, N., Sakamoto, S., Endo, S., Smeets, D., Solder, B., Belohradsky, B. H., Kaloustian, V. M., Oshimura, M., Isomura, M., Nakamura, Y. and Komatsu, K.
TITLE Positional cloning of the gene for Nijmegen breakage syndrome
MEDLINE Nat. Genet. 19 (2), 179-181 (1998)
REFERENCE 2 (bases 1 to 56500)
AUTHORS Matsura, S., Tauchi, H. and Komatsu, K.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1998) Shinya Matsura, Hiroshima University, Research Institute for Radiation Biology and Medicine, Department of Radiation Biology; Kasumi 1-2-3, Minami-Ku, Hiroshima
```

734-8553, Japan (E-mail:shinya@ue.ipc.hiroshima-u.ac.jp,
Tel:81-82-257-5811, Fax:81-82-256-7101)
Sequence updated (26-May-1998).

COMMENT
FEATURES
Source

```
Location/Qualifiers
1..56500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8q21"
/sub_clone="RG255A7"
4790..4848
/number=1
4790..4811
join(4812..4848,6518..6651,7850..7998,8480..8639,
11050..11153,18083..18200,18816..19009,24866..24963,
30519..30648,33818..34090,35682..36129,41481..41549,
43078..43233,46007..46120,52298..52347,53761..53791)
/codon_start=1
/product="NBS1"
/protein_id="BAA28616.1"
/db_xref="GI:3169125"
/translaton="MKLLPAAAGPAGGEPVRLLTGYEVVVGKKNCAILITENDOSISRN
HAVLTANFVSNLTSDTEIPVLTLDKNSKYGTVPNEEKWNGFSRTLKSGDGTTFGVF
GSKFLEYELVACSSCLDYSGKTLNQAIILOLGGFTVNNWTEECTHLYVSVKVTIK
TICALICGRIKPEYTFEFLKAVESKKQPPQIESYPPDLDEPSIGSKNVDLSGRGR
KQIFGKTFIFLNAKOHHKLLSSAVFVGGEARLITEVEEHNFFLAPGTCVYVDGIT
NTTGPSLSQGVSDKLMPSAPVNTTVYADTESQADTWLSEPKETKYSKMEQK
FRLSODATPVRESCKTSSNNNSMVNTLAKWRIPNYQLSPKLPINSKSKORASQOQ
QNSIRNYQPSPTKKRERDEENQESSKSARIETSCSLLEQTPATPSLWKNKEOHL
SENPVDTSIDNNKLFDTDLKSIVKNSASKHAAKLRNKKRKMDDVAIEDEVLSOL
EKEDSLWAKETISNNDKLQDDSEMLPKKILLTEFRSLVINKSTSRNPSINDDYQSL
KNFKFKKVTYPGAGKPLHIGGSDLIAHARKNTELEEWLRQEMEVQNOHAKESLA
DDLFRNPNYLKRER"
4849..6517
/number=1
6518..6651
/number=2
6652..7849
/number=2
7850..7998
/number=3
7999..8479
/number=3
8480..8639
/number=4
8640..11049
/number=4
11050..11153
/number=5
11154..18082
/number=5
18083..18200
/number=6
18201..18815
/number=6
18816..19009
/number=7
19010..24865
/number=7
24866..24963
/number=8
24964..30518
/number=8
30519..30648
/number=9
30649..33817
/number=9
33818..34090
/number=10
34091..35681
/number=10
```

```
exon 35682..36129
      /number=11
intron 36130..41480
      /number=11
exon 41481..41549
      /number=12
intron 41550..43077
      /number=12
exon 43078..43233
      /number=13
intron 43234..46006
      /number=13
exon 46007..46120
      /number=14
intron 46121..52297
      /number=14
exon 52298..52347
      /number=15
intron 52348..53760
      /number=15
exon 53761..53941
      /number=16
3'UTR 53792..53941
polyA_signal 53921..53926
BASE COUNT 16887 a 9935 c 10693 g 18985 t
ORIGIN

Query Match 42.3% Score 1862.8; DB 9; Length 56500;
Best Local Similarity 94.1%; Pred. No. 0;
Matches 2024; Conservative 9; Mismatches 26; Indels 91; Gaps 6;

Qy 2278 tcttttagatcacatccctatttataaaagagagaataaactgaggtatttataaaagaag 2337
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 53752 TCCTTGCAGATACATCCTTATTATAAAGAGAGAGATAACTGAGGATTTTAAAGAAAG 53811

Qy 2338 ccatggaaaaactctctagtagtaagcatctacttcaggccaagaagttatgaatatata 2397
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 53812 CCATGGAAAACTTCCTAGTAGCATCTACTTACGCCCAACAAGGTTATATGAATATATA 53871

Qy 2398 gtgtatagaagcgatttaagttacaatgttttatggcctaatttataataaaatgca 2457
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 53872 GTGTATAGAAGCGATTAAAGTTACAATGTTTATGGCCCTAAATTTATTAAATAAATGCA 53931

Qy 2458 caaaactttgattcttttgtatgtaacaaatgtttgtctgttttcaggctttgtcattg 2517
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 53932 CAAAACITTGATCTTTGTGTATGTAAACAATTTGTTGTTCTGTTTTCAGSCITTTGTCATTG 53991

Qy 2518 catcttttttcatttttaaatgtgttttttttattataaatagtttaatatagcacagttc 2577
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 53992 CATCTTTTTTTCATTTTAAATGTGTTTGTATTATAATAGTTAATATAGTCACAGTTC 54051

Qy 2578 aaattctaaatracgtaaagtaaggaactaaagtaacccctccaccattgtcctagct 2637
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54052 AAAATTCTAAATGACGTAAAGGTAAA-GACTAAAGTCAACCTTCACCATTGTCCTAGCT 54110

Qy 2638 act----- 2640
      |||
Db 54111 ACTTGGTTCCTCAGAAAAAATTCATGATCATCTTCTTATGAATCTTTCCAGGAT 54170

Qy 2641 -----tattttaaatatttcctacacaaatgatagcataac 2678
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54171 TTTTGAGTCCCTATTCAAAATTCCTATTATTTAAATAATTTCTACACAAATGATAGCATAAC 54230

Qy 2679 atatgcagtgcttacacctgtcttttttaacttagtaagattataaaattataggaataac 2738
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54231 ATATGCAGTGTCTACACCTTGTCTTTTACTTAGT-AGATTAAATAATATAGGAATATC 54289

Qy 2739 aatataatgtttttaataatttttttccattatgtcttagctacacctaaactctgg 2798
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54290 AATAATATGTTTTTAATAATATTTTTTCTTTCCATTATGCTGAGTCTTACCTAAACTCTGG 54349

Qy 2799 tgatccaaacaaaatggcttcagtggtgcagatgtcacctacatgttattcttagtactag 2858
```



```

repeat_region /rpt_family="Alu"
7558..7871 /rpt_family="Alu"
7938..8238 /rpt_family="Alu"
9286..9467 /rpt_family="MIR"
9941..10012 /rpt_family="MIR"
11013..11334 /rpt_family="Alu"
11651..11957 /rpt_family="Alu"
12133..12441 /rpt_family="Alu"
12513..12804 /rpt_family="Alu"
12808..12912 /rpt_family="L1"
12913..13213 /rpt_family="Alu"
13214..13638 /rpt_family="L1"
13639..13750 /rpt_family="L2"
14231..14529 /rpt_family="L1"
16063..16506 /rpt_family="Alu"
17682..17990 /rpt_family="Alu"
18183..18704 /rpt_family="ERV"
18757..18921 /rpt_family="MIR"
18939..19063 /rpt_family="Alu"
19103..19178 /rpt_family="MIR"
19312..19448 /rpt_family="MIR"
19626..19961 /rpt_family="L2"
20170..20245 /rpt_family="MER1_type"
20787..20974 /rpt_family="L1"
21556..21787 /rpt_family="L1"
21793..21917 /rpt_family="Alu"
21918..22210 /rpt_family="Alu"
22211..22380 /rpt_family="Alu"
22401..22804 /rpt_family="L1"
22971..23110 /rpt_family="L2"
23617..23881 /rpt_family="Alu"
24264..24633 /rpt_family="MALR"
24697..24908 /rpt_family="MIR"
25116..25179 /rpt_family="L1"
28103..29649 /rpt_family="L1"
29677..31685 /rpt_family="L1"
31689..31818 /rpt_family="Alu"

```

```

repeat_region 31819..32119
Query Match 42.3%; Score 1861.2; DB 9; Length 107549;
Best Local Similarity 94.1%; Pred. No. 0;
Matches 2023; Conservative 9; Mismatches 27; Indels 91; Gaps 6;

Qy 2278 tcttttagatacaatccttattataaaaggagaagataaactagaggatttttaaaagaag 2337
Db 68939 TCTTTGAGATACAAATCCTTATTTAAAAAGGAGAAGATAACTAGAGATTTTAAAAAGAAG 68880
Qy 2338 ccattgaaaaacttccctagtaagcatctacttcaggccaacagggtatataatata 2397
Db 68879 CCATGGAAAAACTTCCCTAGTAGCATCTACTTCAGGCCAACAAAGTTATATGAATATATA 68820
Qy 2398 gtgtatagaagcgatttaagttacaattttttatggcctaataatttataaataaatca 2457
Db 68819 GTGTATAGAAGCGATTAAAGTTACAATGTTTATGGCCCTAAATTTATTAATAAATGCA 68760
Qy 2458 caaaactttgattcttttggatgaacaattgtttgttcttcttcttcttcttcttcttctt 2517
Db 68759 CAAAACCTTTGATTCTTTTGTATGTAACAATGTTTGTCTCTCTTTCAGGCTTTGTCTATTG 68700
Qy 2518 catcttttttttttttttttttttttttttttttttttttttttttttttttttttttt 2577
Db 68699 CATCTTTTTCATTTTAAATGTTTGTATTAATAATAGTTAAATATAGTCACAGTTC 68640
Qy 2578 aaaaattctaaatrtaagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaag 2637
Db 68639 AAAATCTTAATATACGTAAAGTAAA-GACTAAAGTACCCCTTCACCATTTGTCTAGCT 68581
Qy 2638 act----- 2640
Db 68580 ACTTGGTTCCTCCCTCAGAAAAAATTCATGATACATCTTCTTATGAATCTTCCAGGGAT 68521
Qy 2641 -----tatttttaataatttccctacacaaatgatacataac 2678
Db 68520 TTTTGAGTCTCTATTCAAATTCCTATTTTAAATAATTTCTTACACAAATGATAGCATAAC 68461
Qy 2679 atagcagtggtctacacctgttttttacttagtaagataaaaaattataggaataac 2738
Db 68460 ATATGCGAGTGTCTACACCTTGCTTTTACTTAGT-AGATTAAAAATTTATAGGAATATC 68402
Qy 2739 aataataatgtttttaataatttttttttttttttttttttttttttttttttttttttttt 2798
Db 68401 AATAATAATGTTTAAATAATTTTCTTTTCCATTAATGCTGACTCTTACCTTAAACTCTGG 68342
Qy 2799 tgatccaaacaaatggcttcagtggtgcagatgcacacatgatttatttatttatttattt 2858
Db 68341 TGATCCAAACAAAAATGGCTTCAGTGGTGCAGATGTCACCTACATGTTATTTCTAGTACTAG 68282
Qy 2859 aaactgaagaccatgtggagacttcatcaaacatgggttttagttttccaccagaatggaaa 2918
Db 68281 AAAGTGAAGACCATGTGGAGACTTCATCAACATGGGTTTAGTTTTCACCAAGATGGAAA 68222
Qy 2919 gacctgtaccccttttttgggtgttcttactagctgggtgggtgtgtgtgtgtgtgtgtgtgt 2978
Db 68221 GACCTGTACCCCTTTTGGTGGTCTTACTGAGCTGGGTGGGTGCTGCTTTTGGAGCTTAT 68162
Qy 2979 tagagtcctagtttcttacttataaaagtagaagatggtgagattgttttttttttttttt 3038
Db 68161 TAGAGTCTAGTTTCTCTACTTATAAAGTAGAATGGTAGATTGTTTCTTTTCTTCTTACC 68102
Qy 3039 ktaaaagggagatggtaagaacaatgaatgtcttttttcaaacattttatgcagaagtgtatt 3098
Db 68101 TTAAGGGAGATGGTAAGAAACAAATGAATGCTTTTTCAAACTTTTATGACAAAGTGAT 68042
Qy 3099 ttcaagtctgtttcaaaaaatatattcatgtacctgtgatccagaagaagggtgtccca 3158
Db 68041 TTCAAGTCTGTGTCAAAAAATATATTTCATGTACCTGTGATCCAGCAAGAGGGAGTCCCA 67982
Qy 3159 gtcaagagtcactacaactgatttagttgttttagagaatgagaatggaacagtgaagaaat 3218
Db 67981 GTCAAGAGTCACACAACTGATTAGTTGTTTAGAGAAATGAGAAATGGAACAGTGAGGAAT 67922

```


QY 3219 ggaggccatatttcccatgacttcccttgttaacagaagcaacagaagggagagagctg 3278
|||||
Db 67921 GGAGGCCATATTTCCATGACTTCCCTTGTTAAACAGAAGCAACAGAAGGAGCAAGAGCTG 67862

QY 3279 gcctctacatcactctcaactcccaaatcttgtggaagtgcactctacttgcagaaacaa 3338
|||||
Db 67861 GCCTCTACATCACTCTCACCTTCCAATCTTGTGGAAGTGCATCTACTTCCCAAGAACAA 67802

QY 3339 attaaacttacttcaagttctgctctgagtgagtgaaactccagctgcaagagagttag 3398
|||||
Db 67801 ATTAACCTACTTCCAGTTCTGCTGCTTCAGGTGAAGTCACTCAGCTGCAAGGAGTTAG 67742

QY 3399 ggaatgaagtgctttttttaaagctctctcagccttcccttagggaacagaattgggtga 3458
|||||
Db 67741 GGAATGAAGTGCTTTTTTAAAGCTTCTCAGCCTTCCCTAGGAACAGACATTGGGTGA 67682

QY 3459 gccaatctgcaattctactacaggaagcaattgagaccaggttagattattgaatatataga 3518
|||||
Db 67681 GCCAATCTGCAATTTCTACTACAGGCATTGAGACCAGTTAGATTTAGATTTAGAAATATTAGA 67622

QY 3519 gagttatgaacacttaaatattatgagtgtgatgacattgacattgataacatggatacttt 3578
|||||
Db 67621 GAGTTATGAACACTTAAATTTATGATAGTGTGATGACATTTGGATGAGACATGGGATACCTT 67562

QY 3579 agaagtagaattgacaggggcataattagttgatgaaatggagtgactttgagttcctytaa 3638
|||||
Db 67561 AGAAGTAGAATTGACAGGGCATATTAGTTGATGAATGGAGTCATTTGAGTCTCTTAATA 67502

QY 3639 gccatgtatacaattaccagaagtgaagctggtggaacatagtgctccattttacagtta 3698
|||||
Db 67501 GCCATGTATCATTAATTACCAAGTGAAGCTGGTGGACATATGCTCTCCATTTTACAGTTA 67442

QY 3699 aggaataatgacagattaaatattgttctgtcatgccacaatccctttcctaagga 3758
|||||
Db 67441 AGAATATAATGACAGATTAAATTTGTTCTGTCTGTCATGCCCAATCCCTTTCTAAGGA 67382

QY 3759 agactgccctactatagcagtttttataattttgtcaattttatgaatataatgaatgaggag 3818
|||||
Db 67381 AGACTGCCCTACTATAGCAGTTTATATTTGTCAATTTATGAATATAATGAATGA-GAG 67323

QY 3819 ttctgttacctctgtcttttacaataattgggtgtgtgcagattttttccctttttaac 3878
|||||
Db 67322 TTTCTGGTACCTCTGCTCTTTACAAATATTGG- - - TTTGTTCAGTATTTTTCCCTTTTAA 67267

QY 3879 cmttcccaattcgggtgtgtaggtggatgtttccatttgggttttaattgttatccct 3938
||: ||||| ||||| |||||
Db 67266 CCATTCCAATCGGTGTAGTGATG- - - -TTTCATTTTGGTTTAAATTTGTATATCCCT 67212

QY 3939 gatagctataattgggttcataagaaattcctttatcacattctagatgcaagtcctcttgcgg 3998
|||||
Db 67211 GATAGCTATAATTGGGTCTATAGAAATTTCTTTATACATTTCTAGATGCAAGTCTCTTGTGCG 67152

QY 3999 atatacgtattgagatatcacacctagctcgtggctgaactgttttttctatgtcttttg 4058
|||||
Db 67151 ATATACGTATTGAGATATTACACCTAGTCTGTGGCTTTGACTGTTTCTTTTATGTCATTTTG 67092

QY 4059 atgaatagaagttttaaattttgacaaggtcaaatatttttttcttttctttgtttgatt 4118
|||||
Db 67091 ATGAATAGAAGTTTAAATTTTGACAAGGTCAAAATTTATTTTCTTTTGTGTTGATATT 67032

QY 4119 ttttctccaatttaaccccaagatttcagatatctctgctctattatataataaactttata 4178
|||||
Db 67031 TTTTCTCTCCAATTTAAACCCCAAGATTTTCAGATATTCTGCTCTATTATATAAACTTTATA 66972

QY 4179 tttttatttggatcactacactgaattgatattgtattgttgaattatgatacagggttc 4238
|||||
Db 66971 TTTTATATTATTGATCTACTCTTCAATTTGATATGTATGTTGTGTAATTTATGATCAGGGTT 66912

QY 4239 cttttttcccccatacaagttaccagtcattgttaacactgtttattgaagaattatcc 4298
|||||
Db 66911 CTTTTTTTCCCCCATCAAGTATCCAGTCAATGTTTAACACTGTTTATTGAAGAATTAATCC 66852

QY 4299 ttctctcattaaattaccttggccaattagtaaaaaaatcaatcaaccatrm 4348
|||||
Db 66851 TTTCTCTATAAATTACCTTGGCAATTAGTAAAAAATCAATTAAACCATAA 66802

RESULT 6
AK001017
LOCUS
DEFINITION
 Homo sapiens CDNA FLJ10155 fis, clone HEMBA1003433, highly similar
 to Homo sapiens gene for NBS1.
ACCESSION
 AK001017
VERSION
 AK001017.1 GI:7022031
KEYWORDS
 oligo capping; fis (full insert sequence).
SOURCE
 Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to
 mRNA, clone_lib:HEMBAL clone:HEMBA1003433.
ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
 1 (bases 1 to 2044)
AUTHORS
 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
 Nishikawa,T., Nagai,K., Sato,H., Sugano,S., Shiratori,A., Sudo,H.,
 Wagatsuna,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
 Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
 Ono,T., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
 Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
 Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 2044)
REFERENCE
 Isogai,T. and Otsuki,T.
AUTHORS
 Direct Submission
TITLE
 Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genomics@hri.co.jp, Tel.:81-438-52-3951, Fax:81-438-52-3952)
COMMENT
 NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; cDNA full insert
 sequencing: Research Association for Biotechnology; cDNA library
 construction, 5'- & 3'-end one pass sequencing and clone selection:
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.
FEATURES
 Location/Qualifiers
 1. .2044
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEMBA1003433"
 /tissue_type="whole embryo, mainly head"
 /clone_lib="HEMBAL"
 /dev_stage="embryo, 10 weeks"
 /note="cloning vector: pME18SFL3"
 1282^1283
 /note="118 bases segment is present in AF051334 and
 AF058696"
 misc_difference 1378
 /note="compared to AF051334 and AF058696"
 /replace=""
BASE COUNT
 723 a 384 c 442 g 495 t
ORIGIN

Query Match 41.7%; Score 1836.4; DB 9; Length 2044;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 1976; Conservative 0; Mismatches 1; Indels 119; Gaps 2;

QY 13 cgcggttgacgtcgcccgagccctgagagcgagccagctgtggaactgtggaactgtgcccgc 72
|||||
Db 67 CGCGGTTGACGTGCGCGCCCGAGCCCTGAGGAGCGGACCGCATGTGGAACTGCTGCCCGC 126

QY 73 cgcgggccgagcgaggaaccatacagacttttgcgtggcgttgacgtgtgtg 132
|||||
Db 127 CGCGGGCCGCGAGGAGAGACCATACACTTTTGCCTGGCGTTGAGTACGTTGTTGG 186

QY 133 aaggaaaaactgtgccattctaattgaaaaatgatcagtcagtcagcgcaaatcatgctgt 192

Db 187 AAGGAAAACTGTGCCATTCTAATTGAAATGATCAGTCGATCAGCCGAAATCATGCTGT 246
Qy 193 gtaactgtactttctgttaaccaacctgagtcaaacagatgaatccctgtattgac 252
Db 247 GTTAAGTCTGAACCTTTCTGTGAACCAACTGAGTCAACAGATGAATCCCTGTATTGAC 306
Qy 253 attaaagataattctaagtatggtacccttgttaattgaggaataatgcagaatggctt 312
Db 307 ATTAAGAGATAATTCTAAGTATGTGACCTTTTCTTAATGAGGAAAAAATGCAGAAATGGCTT 366
Qy 313 tccccgaacttgaaagtcggggatggtattacttttggagtggttgaagtaattcaag 372
Db 367 TTCCCGAATTTGAAGTCGGGGATGGTATTACTTTTGGAGTGTGGAAATTAATTCAG 426
Qy 373 aatagagtatgagccttgggtgcatgctctcttctgtttagatgctctctgggaaactgc 432
Db 427 AATAGAGTATGAGCCTTTGGTTCATGCTCTCTCTGTTTAGATGCTCTGGGAAACTGC 486
Qy 433 tttaaatcaagctatatgcaacttggaggatttacttgaacaaattggacagaagaatg 492
Db 487 TTTAATCAAGCTATTATGCAACTTTGGAGGATTTACTGTAAACAAATTGGACAGAAGATG 546
Qy 493 cactcacctgtcatggtatcagtgaaagttaccattaaacaatatatgcaactattg 552
Db 547 CACTCACCTTGTCTGATGAGTATCAGTGAAAGTTACCATTAAACAATATATGTGCACCTATTG 606
Qy 553 tggacgtccaattgttaaagccagaaatttttactgaattcctgaaagcagtttcagttccaa 612
Db 607 TGGAGCTCCAAATTGTAAGCCAGAAATATTTACTGAAATTCCTGAAGCAGTTCAGTCCAA 666
Qy 613 gaagcagcctcacaaattgaaagtgttttaaccaccttctgatgaaccatctattggaag 672
Db 667 GAAGCAGCCTCCAAATTTGAAAGTTTATACCCACTTCTGATGAACCAATCTATTGGAAG 726
Qy 673 taaaaattgtatctgtcagacgcaggaagaagaaacaaatcttcaaggaagaaacatt 732
Db 727 TAAAAATGTTGATCTGTCTGACGCGCAGGAAGAAACAAATCTTCAAGGGAACAAAT 786
Qy 733 tatatttttgatgccaacagcataaagaattgagttcccgagttgcttcttgagggtgg 792
Db 787 TATATTTTGAATGCCAAACAGCATAAGAAATTGAGTTCCGACAGTTGCTTTTGGAGGTG 846
Qy 793 ggaagctaggtgtataacagaagagaatgaagaagaacataattcttcttggctccggg 852
Db 847 GGAAGCTAGGTGTATAACAGAAGAGAATGAAGAACAATAATTTCTTTTGGCTCCGGG 906
Qy 853 aacgtgtgtgtgtatagacaggaataacaaactcacagaccttaattctctgactgtcgaa 912
Db 907 AACGTGTGTTGTGATCAGGAATAACAACTCACAGACCTTAATTTCTGTGACTGTGAA 966
Qy 913 gaaatggattcagtcataatggtatgctccaaaggcaaggtcttagacacctattctctga 972
Db 967 GAAATGGATTCACTCAATAATGGATATGCTCCAAAGGCAAGGTCTTAGACCTATTCTCTGA 1026
Qy 973 agcagaatttgatggcggtgattttcatgactacaagaataactgtatgactccagg 1032
Db 1027 AGCAGAAATTTGGATTGGCGGTGATTTTCATGACTACAAAGAAATTTACTGTGTATCTCAGGG 1086
Qy 1033 ccattccagtcagagattaaagacaacaaactccaggaccaagcctttcacaaagcggtgc 1092
Db 1087 CCATCCAGTACAGATTAAAGACACAACTCCAGGACCAAGCCTTTCAACAGGCGGTGC 1146
Qy 1093 agttgatgaaaaataatgccaagcgccccagtgaaacactacaacatcagtagctgacac 1152
Db 1147 AGTTGATGAAAAAATAATGTCGAAGCGGCCAGTGAACACTACAACATACGTAGCTGACAC 1206
Qy 1153 agaatcagacgaacgaatcatggttgaattgaaggccaaggaataacaaagtctc 1212
Db 1207 AGAATCAGAGCAGCATACATCGGGATTGAGTGAAGAGGCCAAAGAAATCAAAAGTCTC 1266
Qy 1213 caaatggacaaaaatttcagaaatgcttttcacaagacgcacccactgtaaaaggagtcctg 1272
|||||

Db 1267 CAAAATGGAAACAAAAA----- 1282
Qy 1273 caaaacaagctctaataataatagtgatggtatcacaatactttggctaaagatgagaatccc 1332
Db 1283 ----- 1282
Qy 1333 aaactatcagctttccaccaactaaattgccaagtataaaataaaagtaaaagtaggcttc 1392
Db 1283 -----CACCAACTAAATTTGCCAAGTATAAATAAAAGTAAAGATAGGCCTTC 1328
Qy 1393 tcagcagcagcagacccaactccatcagaaaactcttctcaagccgtctacc-aaaaaaaggg 1451
Db 1329 TCAGCAGCAGCAGACCAACTCCATCAGAACTACTTTTCAGCCGTCTACCAAAAAAAGGG 1388
Qy 1452 aaagggatgaagaaatcaagaaatgctctcatgcaaatcagcaagaatagaaacgtctc 1511
Db 1389 AAGGGATGAGAAATCAAGNAATGCTTCATGCAAAATCAGCAAGATAGAAACGTCTT 1448
Qy 1512 gtctcttttgaacaaacaaacctgtcacaccctcatttggaaaaataaaggagcagc 1571
Db 1449 GTTCTCTTTTAGAACAAACACAACTGCTACACCCTCATTTGTGGAATAAAGGAGCAGC 1508
Qy 1572 atctatctgagaatgagcctgtggacacaaactcagaacaaataacttattacagatacag 1631
Db 1509 ATCTATCTGAGAAATGAGCCTGTGGACACAAACTCAGACAAATAACTTATTACAGATACAG 1568
Qy 1632 attaaaaatctattgtgaaaaattctgcagtaaatctcatctgcagaaaaagcctaagat 1691
Db 1569 ATTTAAATCTATTGTGAAATTTCTGCCAGTAAATCTCATGCTCAGAAAAGCTTAAGAT 1628
Qy 1692 caaataaaaaagggaatggatgatgtggccatagaagatgaagctatttggaaacagttat 1751
Db 1629 CAAATAAAAAAGGAAATGGATGATGTGGCCATAGAAGATGAAGTATTGGAAACAGTTAT 1688
Qy 1752 tcaaggacacaaacccagagttagaattgtatgtaaaagttcaaaacagaggagaagatg 1811
Db 1689 TCAAGGACACAAACCCAGAGTTAGAAATTTGATGTGAAAGTTTCAAAAACAGGAGGAAGATG 1748
Qy 1812 tcaatgttagaaaaaggccaaggatggatatagaaacaaatgacacttccagttgatgaag 1871
Db 1749 TCAATGTTAGAAAAAGGCCAAGGATGGATATAGAAACAAATGACACTTTTCAGTGATGAAG 1808
Qy 1872 cagtaccagaagtagcaaaaatatctcaagaaaaatgaaattgggaagaaacgtgaactca 1931
Db 1809 CAGTACCAGAAGTAGCAAAAATATCTCAAGAAAATGAAATTTGGGAAGAAACGTGAACATCA 1868
Qy 1932 aggaagactcactgtgtcagctaaagaaatatctcaaatgacaaacttcagatgata 1991
Db 1869 AGGAAGACTCACTATGTCACCTAAAGAAATATCTAACAAATGGCAAACTTCAGGATGATA 1928
Qy 1992 gtgagatgcttccaaaaaagcgtgttattgactgaatttagactcactggtgattaaaaact 2051
Db 1929 GTGAGATGCTTCCAAAAAAGCTGTTATTGACTGAATTTAGATCACTGCTGATTAATAACT 1988
Qy 2052 ctacttcagaaaatccgtctggcataaaatgatgatattatgctgaactaaaaatttc 2107
Db 1989 CTACTTCAGAAATCCGCTGGCATAAATGATGATTATGTTGGTCACTAAAAAATTTTC 2044
|||||

RESULT 7

AF218575 LOCUS AF218575 2605 bp mRNA linear ROD 02-AUG-2000
DEFINITION Rattus norvegicus Nbs1 (NBS1) mRNA, complete cds.
ACCESSION AF218575
VERSION AF218575.1 GI:9651647
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2605)
AUTHORS Lanson,N.A. Jr., Egeland,D.B., Royals,B.A. and Claycomb,W.C.

Db 266 CAGTTTGAGTCAAAACAGATGAATAATTCCTACATTAAACAATAAAAGATAAATTCCTAAGTATGG 325
Qy 277 tacctttgtaatgaggaataaagcagaaatggcgttttcccgaactttgaaagtcggggga 336
Db 326 AACCTTTGTTAATGAAGAAAAAATGCAGACTGGCTCTTCCTGCACGTTGAAGACAGAGA 385
Qy 337 tggattacttttggagtgcttgggaagttaaatcagaatagagatagacgctttggttgc 396
Db 386 TAGAGTTACCTTTTGGGGTCTTTGAAAGTAAATTCAGAGTAGAATACGAGCCCTTGGTTGT 445
Qy 397 atgctctcttcttagatgctctggaataactgctttaaatcaagcttatattgcaact 456
Db 446 TTGTTCTCTGTTAGATGCTCTCGGAAACTGTGTTAAATCAAGCTATTTTACAGCT 505
Qy 457 tggagatttactgtaaacaattggacagaagaatcactcaccttctcatggtatcagt 516
Db 506 TGGAGGACTTACCGCAACAACATGGACGGAAGAAATGTACCCACCTTGTCTATGTCGCAGT 565
Qy 517 gaaagtaccattaaaaaatatgtgcactcatttggagcttccaaattgttaaagccaga 576
Db 566 TAAAGTCAACCATTAACACTATATGTGCACCTCATTTGTGGACGTCCAATTTATAAACACAGA 625
Qy 577 atattttactgaattcctaagacagcttcagtcctcaagaagacagcctccacaattgaaag 636
Db 626 ATACTTTCTGTAATTTCTCAAGCAGTTGAATCTAAGAAACAGCCCTCCAGATATTGAAAG 685
Qy 637 ttttaccacactcttgatgaaccattctattggaagttaaaattgttgactctgcagagc 696
Db 686 TTTTACCACCCATTTGATGAACACGCTATTTGGAAGTAAAGTGTGATCTGTACGGCG 745
Qy 697 gcaggaagaataaaattctcaagggaataacatttatattttgaaatgccaacacaga 756
Db 746 ACATGAAGAAACACAGATCTCAAGAGAAAAACATTTGTGTTTTTAATGCCAACGACGA 805
Qy 757 taagaataatgagttccogagttgcttcttggaggtggggaagctaggttgataacagaaga 816
Db 806 CAAGAAACTCAGCTCGGCGAGTTGCTTTTCGGAGGTGGAGAAGCCAGGCTGTGCGCAAGA 865
Qy 817 gaatgaagaagaacataaattcttttggctcgcggaacgtgtgtgttgatgatacaggaat 876
Db 866 CGACGAAGAGGAACAGAGCTTCTTTTCAGCTCCCGGAACCTGCGTTGTTGATGTAGGAAT 925
Qy 877 aacaaactcacagacaccttaattcctgactgtcagaagaataatggatcagtcataatgga 936
Db 926 AACGAATACACAGCTCATTAATTTCACTCTCCAGAAAAAATGGATTCATTTGATATAATGGA 985
Qy 937 tatgtccaaaggcaaggtccttagaacctattcctgaagcagaaatggattggcggtgat 996
Db 986 TACACTTCAAGGAATGGTCTCAGACCTATTCCTGAAGCGGAGATTGGATTTGGCTGTTAT 1045
Qy 997 tttcatgactacaagaattactgtgatcctcagggccatcccagtcacaggattaaagac 1056
Db 1046 TTTTATGACTACAGAGAATTTACTGTAAATCCGCGAGGCCAGCCTTGTACAGAAATTAAGAC 1105
Qy 1057 aacaaactccaggaaccaagcctttccaagggcgtgtcagttgatgtaaaaaactaatgccaag 1116
Db 1106 AACGACTCCAGGACCAAGCCTTTCCCAAGTCTCTGTCAGCCAAATGGAAAAATTAATCCCAAG 1165
Qy 1117 cgccccagtgaaactacaacatcagtagctgaacagaatcagagaacagcaagcagatacatg 1176
Db 1166 TGCTCCAGTGAATATGACCCACATACGTAGCTGACACAGAATCAGAGCCAGCAGATACATG 1225
Qy 1177 ---ggaatttggatgaaaggccaagaataacaaagtctccaaaattggaacaaaattcag 1233
Db 1226 TATGCTTTTGTAGTGAAGAACACAGAGAAGTAAAGATCCCTGGACTCGAACAAAGCTCTAG 1285
Qy 1234 aatgctttcaagacgaccccaactgttaaaggagtcctgtcaaaacaaagctcttaataataa 1293
Db 1286 GAAACTTTCACAGAACAATTCATATAAAGAGAGGCCCTTAAACCAAGCTCCAAGCTAA 1345
Qy 1294 tagtatggtatcaataactttggctaagatgagaatcccaaaactatcagctttcaccaac 1353
Db 1346 CAACGTAGCATCAGATACGCTGGTTAGAGGAAAGACCCCGAGCTATCAGCTTTTCCCAAT 1405

Qy 1354 taaattggccaagtataataataaagtaaaagatagggtcttcagcagcagcagaccaactc 1413
Db 1406 GAAATTTCTGTTGGCAAAATAAAATAAAGGATTGGACTTCTCAGCAGCAGCAG---AACTC 1462
Qy 1414 catcagaactactttcagccgtctaccacaaaaggggaaggtagaagaataatcaaga 1473
Db 1463 CATCAAAAACCTACTTCCAGCCATGCACAGAAAAGGAAAGGATGAAGACAACACCAGA 1522
Qy 1474 aatgtcttcattgcaaatcagcaagaataagaacgtcttctctcttttagaacaacacaca 1533
Db 1523 GCTGTCTCTGTGCAAAATCATCCAGGATGGAGCTGTCTGTTCCTCTTTTAGAACAAACACA 1582
Qy 1534 acctgtcacccctcattgtggaaaaataaggagcagcatctctctgagatgagcctgt 1593
Db 1583 ACCTCGCGGACCCCTCACTGTGTGAAAAGCAAGG---AGCATCAGTCTCAGAAATCGGACCT 1639
Qy 1594 ggacacaaactcgagacaataaactatttacagatacagatttaaaatctctttagaaaaa 1653
Db 1640 GGACAGGGAAGCGCACACCTCATCTGTGGTGGGATGGACATAGAACTCAACAGGAGAG 1699
Qy 1654 tctgccagttaactctcatgctcagaaaagctaaagctaatcaataaaaaagggaaatgga 1713
Db 1700 TCTGTACAGGAACCACTTCCCACAGAAACTCTTAGACC---CAGAAAAAGAAAAGATGT 1756
Qy 1714 tgaatgtggccatagaagatgaagtatttggacagtttattcaaggacacacaaacccagtt 1773
Db 1757 TGATTTAGCTACAGAGAGAGAACTCTTGGAAAGATTACTCAGGAGTACAAAGCCAGAGTT 1816
Qy 1774 agaaattgatgtgaagatttcaaaaacaggaggaagtgtcaattgttagaaaaagggccaag 1833
Db 1817 GGCAGTTCAAGTTGAAGTTGAAAAGCAGGAGCAGATGACACCATCAGAAAAAAGCCAAAG 1876
Qy 1834 gatggatatagaacaaaatgacactttcagtgatgaagcagtagtaccagaagtagcaaat 1893
Db 1877 GATGGACGAGAGAGGAATCGGCCCTGTAATGGTGGATCAGACCGGAAAAACACACGCC 1936
Qy 1894 atctcaagaaaaatgaaattgggaagaaacgtgaactcaaggaagactcactatggtcagc 1953
Db 1937 TCCTTCAAGAAGATGAAGAGAAAAGAAAGATGAACITTCACACAGAGTCTGTGGTCAACAAA 1996
Qy 1954 taagaataatctacaatagcaaaacttcaggatgtagtagtgatgcttccaaaaaagct 2013
Db 1997 ACATGAAATAGCTTAATAGTGTATGCTTTCAGGACAGCAGTGTGAGGAGCTGCACGGAAC 2056
Qy 2014 gttattgactgaatttagatcactggtgattaaaaactctacttccagaaatcggtctgg 2073
Db 2057 GCTGCTGACTGAATTTTAGGTTCATTTGGTTGTAGTAACCACTCCACACTCCAGAAATCT 2116
Qy 2074 cataatgatgattatggttcaactaaaaaatttcaagaaattccaaaaggtccatatcc 2133
Db 2117 ATGTGTAATGAATGTGGTCCACTGAAGAAATTTCAAGAAATTTCAAGAAGCGCATTTCC 2176
Qy 2134 tggagcagaaaaacttccacataatattggagatcagatctctaataagctcatgctcg 2193
Db 2177 TGGGGCAGGAAACCTTCCACATCATCGGAGGATCAGACTTAGTAGGCCACCATGCTCG 2236
Qy 2194 aaagaatacagaactagaagtggtctaaaggcaggaatggagggtacaaaatcaaatgc 2253
Db 2237 AAAGAATACCGAGTTTAGAAGATGGTTGAACAGAGGAATGGAGGTACAGAAACAACAGC 2296
Qy 2254 aaaagaagagctctctgtgatgatcttttttagatacaactccttatttaaaagggagaag 2313
Db 2297 AAAGGAGGAATCTCTTGTGATGATCTGTTTAGATATAATATCCTTAATGTAATA---AGAAG 2353
Qy 2314 ataactgaggaattttaaaaagaagccatggaaaaacttctctagtagtaagcatctacttcag 2373
Db 2354 ATAAATTGAACCTTTGTAAGGAAGCTGTAGAAACACATTTTCTACTTTCAGACCAACTGAGAT 2413
Qy 2374 ccaacaaggttatatgaatatatagtgat 2403
Db 2414 ACATTATAATAGTACATAGAAATTTATAAAT 2443

RESULT 9
AF092840
LOCUS
DEFINITION Mus musculus nibrin mRNA, complete cds.
ACCESSION AF092840
VERSION AF092840.1 GI:3676834
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2473)
Mas,C., Bourgeois,F. and Simonneau,M.
Isolation of 50 cDNAs differentially expressed in embryonic
forebrain as compared to mid and hindbrain : a strategy to identify
candidate genes involved in human neurodevelopmental diseases
Unpublished
2 (bases 1 to 2473)
Mas,C., Bourgeois,F. and Simonneau,M.
Direct Submission
Submitted (16-SEP-1998) Neurogenetique, INSERM CRI 9701, Hopital
Robert Debre 48 Bvd Serurier, Paris 75019, France
FEATURES
source
1..2473
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="brainstem"
/dev_stage="neonate"
91..2346
/codon_start=1
/product="nibrin"
/protein_id="AAC62113.1"
/db_xref="GI:3676835"
/translation="MNKLLPAAGASLGEPLRLAGVEYVYVGRKNCGILINDOSISRN
HAYLVNFPVTSLSQDEIPTLTIKNSKYGVTPVNEKMTGLSCTLKIGDRVTCGVF
ESKFRVEPLVVCSSCLDVSKTVLQALLOGLTANNWTECHLVNYSVKVTK
TICALCGRIKPEYFELKAVESKQPPDIETFPIDFAIGSKSVLDLGRH
KQIFKGTIVFLNAKHKLLSSAVAFGGAEALMAEDDEEQSFSAFGTCVVDVGIT
NTQLIISHQKWIHLIMDTLQNRGLRPIPEAIEGLAVIFMTNFCNPGQPCTELK
TTPGPSLSQVLSANGKIIIPSPVNNMTYVADTSEPADTICMPLSERPEVKIPGLEQ
SSRLKSOEYFNTEAKPKSKANNVADTLVRGKTPSYOLSPMKPVVANKNDWTSQ
QONSINKYPOCTKRDEREDNPELSCKSRMBELSCSLLEQTPAGPSLWKSKEHQ
QNTLDRADTSVSGMDIELNRKSPDRPLPTETLRPRKRDVDIATSEEVLEELR
SKPELAVQVKEQBADDITRRKPRMDAERNRNLNGSEPSNSALQDEDEKRLR
QTESWSTKEIANSGLQDSSELPRLKLLTFEFLSVNHNHNSRNLCVNECGPLKN
FKFKKATFPGAKPLHIIIGGSLVGHARKNTELEWLKQEMEVQKQAKESLADD
LFRYPNVKRR"

BASE COUNT 829 a 509 c 575 g 560 t

Query Match 31.8%; Score 1398.8; DB 10; Length 2473;
Best Local Similarity 75.9%; Pred. No. 4.3e-241;
Matches 1798; Conservative 0; Mismatches 557; Indels 15; Gaps 5;

QY 37 ctgagggccggaccgagtgygaactgctgcccgcgcccggccggcaggagagagaacc 96
DB 75 CTGAGAGCGCGGCCCATGTGGAAGCTACTCCCGCGCGCGGTGCGTCACTAGGAGAACC 134
QY 97 atcacagacttttactgctgttagtacgtgtgttggaaggagaaactgtgccattcta 156
DB 135 ATACCGACTTTTGGCGCGGTGAGTACGTGTGTGGAGAGAAAACGTGGCATTCGTAT 194
QY 157 tgaatgatcagtcagtcagcgcgaatacgtctgtgttaactgtaacttttctgtaac 216
DB 195 TGAATGATGATCACTCACTCGAACCAATGCTGTCTTAACAGTAACATTCCTCTGTAC 254
QY 217 caacctgagtcacacagatgaatccctgtatttgacattaaaagataattcttaagtatgg 276
DB 255 CAGTTTGAGTCAACAGATGAATTCCTACATTAACAATAAAGATAAATCTTAAGTATGG 314
QY 277 tacctttgttaagagagaaaaaatgcagaatggcgttttcccgcaacttgaagtcgggga 336

DB 315 AACCTTTGTTAATGAAGAAAAAATGCAGACTGGTCTTTCTCTGCAGCTTGAAGACAGSAGA 374
QY 337 tggattacttttgagtggttgaagtataattcagaataagatagatagcccttgggtgc 396
DB 375 TAGAGTTACCTTTGGGGTGTGTTGAAAGTAAATTCAGAGTAGAATACGAGCCCTTGGTGT 434
QY 397 atgctctcttctttagatgctctgggaaactgctttaaatacaagctatatatgcaact 456
DB 435 TTGTTCTTCTGTTGTAGATGCTCTGGGAAACTGTTTAAATCAAGCTATTATTACAGCT 494
QY 457 tggaggatttactgtaacaatttggacagaagaatgcactcaccttgcctggtatcagt 516
DB 495 TGGAGGACTTTACCGCAACAACTGCAGCGGAAATGTACCCACCTTGTCTATGTGCGGAGT 554
QY 517 gaaagtaccattaaaaaatatgctcactcatttggagcgtccaattgttaagccaga 576
DB 555 TAAAGTCACCATTAATACTATATGTGCACATCTTTTGGAGCTCAATATATAAACACAGA 614
QY 577 atatttactgaattcctgaaagcagttcagtcgaagaagcagcctccacaataatgaaag 636
DB 615 ATACTTTTCTGAATTTCTCAAGCAGTTGAATCTTAAGAAACAGCTCCAGATATTGAAAG 674
QY 637 ttttaccacaccttgaataaacattctattggaagtataaataatgttgcaggacg 696
DB 675 TTTTACCACCCCAATGATGAACAGCTATTGGAAGTAAAGTGTGTGCTGTGTCAGGCG 734
QY 697 gcaggaaagaaaaacaaatcttcaagggaaacatttatatttttgaatgccaacagca 756
DB 735 ACATGAAGAAAAACAGATCTTCAAGGAAACAACTTTGTGTTTAAATGCCAAGCAGA 794
QY 757 taagaaattgagttccgagttgcttggagtggtgggaaagctagggtgataacagaaga 816
DB 795 CAAGAAACTCAGCTCGCAGTTGCTTTCGGAGTGGAGAGCCAGGCTGATGGCAGAAGA 854
QY 817 gaatgaagaagaacataattcttttggctccgggaacgtgtgttcttgcatacagaat 876
DB 855 CGAGGAAGGAAGACAGAGCTCTTTTTCAGCTCCCGGAACCTTGCCTTGTGTGATGAGGA 914
QY 877 acaaaactcagacaccttaattcctgactgctcagaagaaatggattcagtcataatgga 936
DB 915 AACGAATACACAGCTCATATTTTCACACTCCACAGAAAAATGGATTTCATTTGATAATGA 974
QY 937 tatgctccaaaggaaggttcttagacctattcctgaagcagaaatggattggtggtgat 996
DB 975 TACACTTCAAGGAATGCTCTCAGACCTATTTCCTGAAGCGGAGATTTGGATTGGCTGTAT 1034
QY 997 tticagactacaagaattactgtgactcctcagggccatcccgatcacaggattaaagac 1056
DB 1035 TTTTATGACTACAGAGAAATTAAGTAAATCCGAGCGGCGGCTTGTACAGAAATTAAGAC 1094
QY 1057 acaaaactccaggaccacgtcttcaagggctgtcagttgataaaaaactaatgccaag 1116
DB 1095 AAGACTCCAGGACCAAGCTTTCCCAAGTCTCTGTGACGCCAATGGAANAATAATCCCAAG 1154
QY 1117 cgcccagtgaaacactacaacatcagctagctgacacagaatcagagcaagcagatatacg 1176
DB 1155 TGCTCCAGTGAATATGACCACATACGTAGCTGACACAGAATACAGAGCCAGCAGATACATG 1214
QY 1177 ---ggatttgagtgaaaggccaaagaaaatacaagctcctcaaaatggaaacaaattcag 1233
DB 1215 TATGCCCTTTTCAGTGAAGAGCAGAGAAAGTAAAGATCCCTGGACTGGAACAAAGCTTAG 1274
QY 1234 aatgcttccaaagacgcaccctgtaaaggagctctgcacaaacagcctctaataata 1293
DB 1275 GAACHTTTCACAGAAACATTCATATATAAGGAGGCCCTTAAACCAAGCTTCCAAAGCTAA 1334
QY 1294 tagtatggtatcaatactttggttaagatgagaatccccaaactatcagcttttcccaac 1353
DB 1335 CAACGTAGCATGATGATGCTGGTTAGAGGAAAGACCCCGAGCTTATCAGCTTCTTCCAAAT 1394
QY 1354 taattgccaagtataataaaaaagtaaggtcttctcagcagcagaccactc 1413

Db 194 CAGTTTGGTCAACACAGATGAATTCCTACATTAACAAATAAAAGATAATTCCTAAGTAGG 253
Qy 277 tacotttggtaaatgagaaaaatgcagaatggctttcccgaaactttggaagtgggga 336
Db 254 AACCTTTGTTAATGAAGAAAAATGCAGACTGGTCTTCTTCGACGTTGAAGACAGGAGA 313
Qy 337 tggattacttttggagtgtttggaagttaattcagaatagagatatgagcctttggttgc 396
Db 314 TAGAGTTACCTTTGGGGTGTGTAAGTAAATTTCAGAGTAGAATACGAGCCCTTGGTGT 373
Qy 397 atgcctcttctttagatgctctggaaaaactgctttaaactcaagctatatattgcaact 456
Db 374 TTGTTCTCTCTGTTAGATGCTCTGGAAAAACHTGTTTAAATCAAGTATTTTACAGCT 433
Qy 457 tggaggatttactgtaaacattggacagaagaatgcactcacctgtcatgggtatcagt 516
Db 434 TGGAGGACTTACCCCAAAACAACTGGACGGGAAGATGTACCCACTTGTCTCATGTCGGCAGT 493
Qy 517 gaaagttaccattaaaaaatatgtgcactcattgtggacgtccaattgttaagccaga 576
Db 494 TAAAGTCAACCATTAATAACTATATGTGCACATCAATTTGTGGACGTCCAATTATATAAACCCAGA 553
Qy 577 atatttactgaattcctgaaagcagtttcagtcacaagaagcagcctccacaaattgaaag 636
Db 554 ATACTTTTCTGAATTTCTCAAGCAGTTGAATTTTAAGAAACAGCCTCCAGATATTGAAG 613
Qy 637 ttttaccacaccttggatgaaccatctattggaagttaaaatgttgatctgtcaggagc 696
Db 614 TTTTATCCCAACCATTCATGAACCACTATTGGAAGTTAAAGTGTTCATCTGTCAAGGCG 673
Qy 697 gcagggaagaaacaaatcttcaagggaaacaaatktatatttttgaaatgccaaacagca 756
Db 674 ACATGAAGAAAAACAGATCTTCAAGGAAAAACAACTTTGTGTGTTTAAATGCCAAGCAGCA 733
Qy 757 taagaaattagttcccgactgtcttggaggtggggaagctaggtgataacagaga 816
Db 734 CAAGAACTCAGCTCGGCAGTTGCTTTTCGGAGGTGGAGAGCCAGGCTGATGGCAGAGA 793
Qy 817 gaaTgaagaagaacataattcttttggctccgggaacgtgtgtgttggatcacaggaat 876
Db 794 CGAGGAAGGGAACAGAGCTCTTTTCAGCTCCCGGAACCTGCGTGTGTTGATGTAGGAAT 853
Qy 877 acaaaactcacagaccttaattcctgactgcagagaagaatggattcagtcocataatggga 936
Db 854 AACGAATACACAGCTCATATAATTTCCACACTCCCAAGAAAAATGGATTCAATTTGATAATGGA 913
Qy 937 tatgtccaaagggaaggtcttagacctattcctggaagcagaaattgagttggcgtgat 996
Db 914 TACACTTCAAGGNAATGGTCTCAGACCTATTCCTGGAAGCGGAGATTGGATTGGCTGTAT 973
Qy 997 ttctgactacaaagaattactgtgatcctcagggccatccccagtcacaggaattaaagac 1056
Db 974 TTTTATGACTACAGAGAATTACTGTAATCCGACGGCCAGCCTTTGTACAGAAATTAAGAC 1033
Qy 1057 acaactccaggaccagcctttcacaggcgtgtcagttgatgaaaaactaatggccaag 1116
Db 1034 AACGACTCCAAGGACCAAGCCTTTCCCAAGTCTCTCAGCCCAATGGAATAATAATCCCAAG 1093
Qy 1117 cgccccagtgaaacactacacatacgtagctgacacagaatcagagcaagacagatacatg 1176
Db 1094 TGCTCCAGTGANATGACCACATACGTAGCTGAGACAGAAATCAGACGACGACAGATCATG 1153
Qy 1177 ---ggattttagtgaaggccaaagaatacaagttcccaaatggaaacaaatttcag 1233
Db 1154 TATGCCCTTTGAGTGAAGACCAAGGAAGTAAAGATCCCTGGACTGGCAACAAAGCTCTAG 1213
Qy 1234 aatgctttcaagagcagccactgtaaaggagtcctgcacaaacagctctataataa 1293
Db 1214 GAAACTTTTCAAGAAACATTTCAATATTAAGAGGAGGCCCTTAAACCAAGCTCCAAAGCTAA 1273
Qy 1294 tagtgggtacaaacttctggctgaagatgagaatccccaaactcagctttcaccaac 1353
Db 1274 CAAGGTAGCATCAGATACGTGGTTAGAGGAAGACCCCGAGCTATCAGCTTCTCCAAAT 1333

Qy 1354 taaattgccagtgataataaaaaagtaaaagtagggcttctcagcagcagcagacaaactc 1413
Db 1334 GAAATTTCTTGTCCAAATAAAAAATAGGATTGGACTTCTCAGCAGCAGCAG---AACTC 1390
Qy 1414 catcagaaaactactttcagccgtctacccaaaaaagggaaggatgaagaaatacaaga 1473
Db 1391 CTTCAAAAACCTTACTTCCAGCCATGCACGAAAAAGGGAAGGATGAAGACACCCAGA 1450
Qy 1474 aatgtcttcagcaaatcagcaagaatagaaacgtctgttctctcttttagacaacaaaca 1533
Db 1451 GCTGCTCTCGTGCAAAATCATCCAGATGGAGCTGTCTCTGTCTCTTTTAGAACAAACACA 1510
Qy 1534 acctgtcacacctcattgtggaataataagagcagcatcta tctgagaatgagcctgt 1593
Db 1511 ACCTGCCGAGCCCTCACTGTGGAAGCAAG---AGCATCAGTCTCAGAAATGGACCCCT 1567
Qy 1594 ggacaaaaactcagacaataaacttatttacagatcacagatttaaaatctatttgaaaaa 1653
Db 1568 GGACGGGAAGCCGACACCTCATCTGTGGTGGGATGGACATGAACCTCAACAGAGAGAG 1627
Qy 1654 tctgccagtaaatctcatgctgcagaaaagctaagatcaaatataaaaaagggaatgga 1713
Db 1628 TCCTGACAGGAACCACTTCCCACAGAAACTCTTTAGACC---CAGAAAAAAGAAAGATGT 1684
Qy 1714 tgatgtggccataagaagatgaagtattggaacagttattcaaggagcacacaaaccagagt 1773
Db 1685 TGATTTAGCTACACAGAGGAAGTCTTGGAAAGAGTTACTGAGGAGTACAAAGCCAGAGTT 1744
Qy 1774 agaaattgagtgaagttccaaaaacagggaagatgtaattgttagaaaaagccaag 1833
Db 1745 GCGAGTTCAAGTGAAGGTTGAAAAAGCAGGAGCGACGACACCATCAGAAAAAGCCCAAG 1804
Qy 1834 gatgatatagaacaaatgacactttcagtgatgaagcagtcaccagaagtagcaaaat 1893
Db 1805 GATGGACGAGAGAGGAATCGGCCCTGGAATGGTGGATCAGAGCGCGGAAAGCAACAGCGC 1864
Qy 1894 atctcaagaaaaataaattgggaagaacgtgaactcaaggaagactcactatggtcagc 1953
Db 1865 TCITCAAGAAGATGAAGAGAAAAAGAAAGATGAACTTTCAGACAGAGTCGTGGTCAACAAA 1924
Qy 1954 taagaataatctaaacatgacaaactccaggtatgtagtgagatgcttccaaaaaagct 2013
Db 1925 ACATGAAATAGCTAATAGTGTCTTCAGGACAGCAGTGAGGAGGCTGCCACGGAACCT 1984
Qy 2014 gtattgactgaatttagatcacgtgtgattaaaaactctacttccagaaaatccgtctcg 2073
Db 1985 GCTGCTCAGCTGAATTTAGGTCATTTGGTTGTTCAGTAACCACTCCACCTCCAGAAATCT 2044
Qy 2074 cataaTgatgattatggtccaactaaaaatttcaagaaattcaaaagggtcacatatcc 2133
Db 2045 ATGTGTAATGAATGTGTCACCTGAAGAAATTTCAAGAAATTTCAAGAAAGGCGACATTTCC 2104
Qy 2134 tgaagcagggaaaaacttccacacatcattggaggatcagataatagctcatgctcg 2193
Db 2105 TGGGGCAGGAAAGCTTCCACACATCATCGGAGGATGAGACTTAGTAGGCCACCATGTCTG 2164
Qy 2194 aagaatcacagaactagaagatggctcaaggcaggaagtggagtgacaaaatcaactgc 2253
Db 2165 AAAGAATACCGATTAGAAGAAATGTTGAACACAGAAATGGAGGTGAGAAACAAACAAGC 2224
Qy 2254 aagaagagatctcttctgtgatgatcttttttagatacaatccttattttaaaaggagaag 2313
Db 2225 AAAGGAGGAATCTCTTGTCTGATGATCTGTTTAGATATAATCTTAATGTAAAA---AGAAG 2281
Qy 2314 ataactgagattttaaaaaagaagccatggaaaaacttcttagtaagcatctacttcagg 2373
Db 2282 ATAACTGAACCTTGTAAAAGGAAGCTGTAGAAACACATTTCTACTTTCAGACCACTGAGAT 2341
Qy 2374 ccaacaaggttatgaatatatagtgtat 2403
Db 2342 ACATTATATAGTACATAGAAATTTATAAT 2371

| | | | | | |
|-----------------------|--|---------------------|----------------|---------------|---------|
| * 30247 | gap of unknown length | | | | |
| * 30973: | contig of 727 bp in length | | | | |
| * 30974 | gap of unknown length | | | | |
| * 31721: | contig of 748 bp in length | | | | |
| * 31722 | gap of unknown length | | | | |
| * 32456: | contig of 735 bp in length | | | | |
| * 32457 | gap of unknown length | | | | |
| * 33185: | contig of 729 bp in length | | | | |
| * 33186 | gap of unknown length | | | | |
| * 33923: | contig of 738 bp in length | | | | |
| * 33924 | gap of unknown length | | | | |
| * 34648: | contig of 725 bp in length | | | | |
| * 34649 | gap of unknown length | | | | |
| * 35376: | contig of 728 bp in length | | | | |
| * 35377 | gap of unknown length | | | | |
| * 36107: | contig of 731 bp in length | | | | |
| * 36108 | gap of unknown length | | | | |
| * 36832: | contig of 725 bp in length | | | | |
| * 36833 | gap of unknown length | | | | |
| * 37589: | contig of 757 bp in length | | | | |
| * 37590 | gap of unknown length | | | | |
| * 38342: | contig of 753 bp in length | | | | |
| * 38343 | gap of unknown length | | | | |
| * 39092: | contig of 750 bp in length | | | | |
| * 39093 | gap of unknown length | | | | |
| * 39838: | contig of 746 bp in length | | | | |
| * 39839 | gap of unknown length | | | | |
| * 40586: | contig of 748 bp in length | | | | |
| * 41330: | contig of 744 bp in length | | | | |
| * 41331 | gap of unknown length | | | | |
| * 42108: | contig of 778 bp in length | | | | |
| * 42824: | contig of 716 bp in length | | | | |
| * 42825 | gap of unknown length | | | | |
| * 43539: | contig of 715 bp in length | | | | |
| * 43540 | gap of unknown length | | | | |
| * 44280: | contig of 741 bp in length | | | | |
| * 44281 | gap of unknown length | | | | |
| * 45022: | contig of 742 bp in length | | | | |
| * 45023 | gap of unknown length | | | | |
| * 45775: | contig of 753 bp in length | | | | |
| * 45776 | gap of unknown length | | | | |
| * 46518: | contig of 743 bp in length | | | | |
| * 46519 | gap of unknown length | | | | |
| * 47266: | contig of 748 bp in length | | | | |
| * 47267 | gap of unknown length | | | | |
| * 48011: | contig of 745 bp in length | | | | |
| * 48012 | gap of unknown length | | | | |
| * 48758: | contig of 747 bp in length | | | | |
| * 48759 | gap of unknown length | | | | |
| * 49505: | contig of 747 bp in length | | | | |
| * 49506 | gap of unknown length | | | | |
| * 50253: | contig of 748 bp in length | | | | |
| * 50254 | gap of unknown length | | | | |
| * 50998: | contig of 745 bp in length | | | | |
| * 50999 | gap of unknown length | | | | |
| Query Match | 15.7%; | Score 692.8; | DB 2; | Length 64310; | |
| Best Local Similarity | 96.2%; | Pred. No. 1.2e-114; | | | |
| Matches 717; | Conservative | 1; | Mismatches 26; | Indels 1; | Gaps 1; |
| Qy 2774 | tgctgtagtcttaactaaactctggtgatccaacaaatggcttcagtggtgcagatgt | 2833 | | | |
| Db 60577 | TGCTGAATCGCGCTCTTACTACTGTANAGNANACACAATGGCTTCAGTGGTGAGATGT | 60636 | | | |
| Qy 2834 | cacctacatgttacttagtagtaagaactgaagaccatggagacttcatacaaatg | 2893 | | | |
| Db 60637 | CACCTACATGTATTCCTAGTACTAGAACTGAAGA-CATGTGGAGACTTCATCAACATG | 60695 | | | |
| Qy 2894 | ggtttagtttcaccagaaatggaagaccctgtaccctcttttgggtctactgagctg | 2953 | | | |
| Db 60696 | GGTTTAGTTTCCACAGAAAGGAGACCTGTACCCCTTTTGGTGTCTTACTAGACTG | 60755 | | | |

| | | | | | |
|------------|---|----------------------------|------|----------|-----------------|
| Qy 2954 | ggtgggtgtctgtttttagagcttatttagagtccttagtttctacttataaagtgaagt | 3013 | | | |
| Db 60756 | GGTGGGTGTCTGTTTGGAGCTTATTTAGAGTCCTAGTCTTCTACTTATAAGTAGAAAT | 60815 | | | |
| Qy 3014 | ggtgagattgtttctcttttttttacccktaaaagggagatggttaagaaacaataatgctcttt | 3073 | | | |
| Db 60816 | GGTGAGATTGTTTCTTTTCTTACCTTAAGGGAGATGGTAAGAAACAATGAATGCTCTTT | 60875 | | | |
| Qy 3074 | tttcaaacctttatgacaagtgtatttcaagctctgtgttcaaaaatatatcatgtacct | 3133 | | | |
| Db 60876 | TTTCAAACTTTATTGACAAAGTGATTTTCAAGTCTGTGTTCAAAAATATATTTCATGTACT | 60935 | | | |
| Qy 3134 | gtgatccagcaagaaggagttccagtcgaagagtcactacactgattagttggttagag | 3193 | | | |
| Db 60936 | GTGATCCAGCAAGAAGGGAGTTCAGTCAAGAGTCACATGATTAGTTGTTTACAG | 60995 | | | |
| Qy 3194 | aatgagaaatggaacagtgaagaaatggaggccatatattccatgactccctctgtaaacag | 3253 | | | |
| Db 60996 | AATGAGAAATGGAACAGTGAAGAAATGGAGGCATATTTTCCATGACTTCCCTTGTAAACAG | 61055 | | | |
| Qy 3254 | aagcaacagaaaggacagagctggcctctacatcactctcacccttccaaatctgtgg | 3313 | | | |
| Db 61056 | AAGCAACAGAAAGGACAGAGGCTGGCCTCTACATCCTCTCACCTTCCAAATCTTGTGG | 61115 | | | |
| Qy 3314 | aagtgcattctacttcgcagaacaaattaaacttacttccaagtctctggtgctggtgcaggt | 3373 | | | |
| Db 61116 | AAGTGCATCTACTTCCAGAACCAAAATTAACCTTACTTCCAAAGTCTCTGGCTGCTGCAGGT | 61175 | | | |
| Qy 3374 | ggaactccagctgcaaggaggttagggaaatgaaggctcttttttaaaagctcttcagcc | 3433 | | | |
| Db 61176 | GGAACCTCCAGCTGCAAGGGAGTTAGGGAAATGAAGGCTCTTTTAAAGGTTCTTCAGCC | 61235 | | | |
| Qy 3434 | ttcttagggaaacagaaatgggtgagccaatctgcaatctctactacagcattgagacc | 3493 | | | |
| Db 61236 | TTCTTAGGGAAACAGACATTGGGTGAGCCCATCTGGCATTTCTACTACAGGATTGAGACC | 61295 | | | |
| Qy 3494 | agttagattattgaaatattataga | 3518 | | | |
| Db 61296 | AGTTAGATTATTGAAATATTATAGA | 61320 | | | |
| RESULT 12 | | | | | |
| AF230342 | AF230342 | 2533 bp | mrna | linear | VRT 09-JAN-2001 |
| LOCUS | Gallus gallus Nijmegen | breakage syndrome 1 (NBS1) | mrna | complete | |
| DEFINITION | cds. | | | | |
| ACCESSION | AF230342 | | | | |
| VERSION | AF230342.1 | GI:12056575 | | | |
| KEYWORDS | chicken. | | | | |
| SOURCE | Gallus gallus | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| REFERENCE | Archosauria; Aves; Neognathae; Galliformes; Phasianidae; | | | | |
| AUTHORS | Phasianinae; Gallus. | | | | |
| TITLE | 1 (bases 1 to 2533) | | | | |
| JOURNAL | Tauchi,H., Kobayashi,J., Morishima,Ki., Matsuura,S., Nakamura,A., | | | | |
| PUBMED | Shiraishi,T., Ito,E., Masnada,D., Delia,D. and Komatsu,K. | | | | |
| REFERENCE | The Forkhead-associated Domain of NBS1 Is Essential for Nuclear | | | | |
| AUTHORS | Foci Formation after Irradiation but Not Essential for | | | | |
| TITLE | hRAD50.hMRE11.NBS1 Complex DNA Repair Activity | | | | |
| JOURNAL | J. Biol. Chem. 276 (1), 12-15 (2001) | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1. .2533 | | | | |
| | /organism="Gallus gallus" | | | | |
| | /db_xref="taxon:9031" | | | | |

Db 1492 TACAAATTACTTCATCTAGCTAGG-----AAAAGGAAAGAGCTGAAGAGGAGAGAAAC 1548
Qy 1477 gtcttcgcaaatcagcaagaatag--aaacgtcttcttctctttagaacaacacaca 1533
Db 1549 ATCTCTATCCAAACAGCAAACTGGAGAAAAAGCCATTGCCTGTTTCCGAGTCCACCGA 1608
Qy 1534 acctgtcacacctctatgttgaaaaataaggagcagcatctatctgagaatgagcctgt 1593
Db 1609 ATCTCAGCTTCATCAGTCAGTAAGTGTGAAAAAGAACAGC-----ATGGAAGGGAATA 1664
Qy 1594 ggacacaaactcagacaataacttatttacagatacagatttaaaatctatttgaaaaa 1653
Db 1665 ACATCCAGCTTGGAGGAAAGTGGTGAAGTGGCAAGTGAATAAACTGACATTAAATTA 1724
Qy 1654 ttctgccagtaaatctctatcgtcgcagaaaagctaaagtaacaaataaaaaagggaatgga 1713
Db 1725 CTTTTAGTGAATC-----CTGCACCAAGAAAGAAAGAGTGTAGATGATGTCTGA 1779
Qy 1714 tgatgtggccatagaagatgaagtatttggaacagttatttcaaggacacacaaacccagatt 1773
Db 1780 AGATGTAGAGACGCTAGAGATGTTGTTGAAAGCAGAGACCTAGACTGGGAAGGCAAC 1839
Qy 1774 agaattgatgtgaa-----agttcaaaaaacagaggaagatgtcaaatgttagaaaa 1825
Db 1840 AGCAAAATGGTGACCAAGGAAGCTCAAAACCAATAAAACGAAGAAGAGATGCTCGGAACCAA 1899
Qy 1826 aggccaaaggtggratagaaaacaaatgacactttcagttgatgaagcagtacccagaagt 1885
Db 1900 AGGCAGCAAGCAAGCAAGCAATCAAAAGCAGAGAGAGAGAAATCAATGCTTGAGAAA 1959
Qy 1886 agcaaaatatctcaagaaaaatgaaattggga-agaaaactggaactcaagaaagactcaact 1944
Db 1960 AGAAGAAGTTGGATCAGTTCTGACTCTGGAAGACAAAAGCAAGATCAAGGAAGATCATC 2019
Qy 1945 atggtc---agctaaagaataatacaaatgacaaacttcaggatgagatgagatgct 2001
Db 2020 AGTCTCGATAAGAAACAAATTAATTAACCAACAAACAACTTGAAGATGACTCCAGCCGTCT 2079
Qy 2002 tccaaaaagctgttattgactgaatttagactcactggtgatttaaaactctacttccag 2061
Db 2080 TCCTAGCAAACTTCTGTTGACTGAGTTTCGATCAGTCTGTTCTCAGCTGCCAAGATCAA 2139
Qy 2062 aaatccctctggcctaataatgatgattatgttcaactaaaaaatttcaagaaattcaaaaa 2121
Db 2140 CAGCCCAACTATGAGAAATACCAAAATCAGGGGAGAGAAACAAATTTCAAAAGCTTCAGAAA 2199
Qy 2122 ggtcacatatcttgagcagaaaactccacacatcatcttgagatgagatcgaatcgaatgc 2181
Db 2200 GGTCCCTTACCCAGGGGAGACAACTTCCATACATAAATTTGGAGGATCAGATTTGGTTGC 2259
Qy 2182 tcatctgctcgaagaatacagaaactagagagtggttaaggcagggaatggaggatca 2241
Db 2260 TCACCAAGCTAGAAAGAAATTCAGAACTAGAAAGTGTTAAGAGAAAGATTTGGAGGACGA 2319
Qy 2242 aaatcaacatgcaaaagaagctctcttgatgagatgatttttttagatcacaaatccttatt 2301
Db 2320 GAATTCGAGCTGCAAGAGAAAGAAATCGCTGTGTGATGATCTCTTTAGATATGATCCTAATGT 2379
Qy 2302 aaaaagagaagataaactgaggtattttaaagaagccat 2341
Db 2380 GAAGAGGAAGATAAATTTGTGACCTGAAATTACAGCAAT 2419

RESULT 13
AX192776/c
LOCUS AX192776 646 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 343 from Patent WO0149716.
ACCESSION AX192776
VERSION AX192776.1 GI:15210732
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 646)
AUTHORS Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,
Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.
TITLE Compounds for immunotherapy and diagnosis of colon cancer and
methods for their use
JOURNAL Patent: WO 0149716-A 343 12-JUL-2001;
CORIXA CORPORATION (US)
FEATURES Location/Qualifiers
source 1..646
/organism="Homo sapiens"
BASE COUNT 239 a 124 c 95 g 188 t
ORIGIN

Query Match 12.5%; Score 552.2; DB 6; Length 646;
Best Local Similarity 93.0%; Pred. No. 2.9e-89;
Matches 609; Conservative 4; Mismatches 32; Indels 10; Gaps 3;

Qy 3419 aaaagcttcagccttccttagggaacagaaaattgggtgagcccaatctgcaatttctact 3478
Db 645 AAAGCTTTCAGCCTTCCTAGGGAACAGAAATTTGGGAGGCCCAATTCGCAATTTCTACT 586
Qy 3479 acagggcattgagaccagtttagatttgaataattatagagagttatgaacactaaatt 3538
Db 585 ACAGGCATTTGAGACCAGTTAGATTATTGAAATATTATAGAGAGTTATGAACACTTAAATT 526
Qy 3539 atgataagtgatagaacttggttagaacaatgggatacttttagaagtagaattgcagggc 3598
Db 525 ATGATAGTGTATGACATTTGGATAGAACATGGGATACCTTTAGAAAGTGAATTTGACGGC 466
Qy 3599 atattagtgatgaatggagtcatttgagtcctytaaagccatgatacaataacc 3658
Db 465 ATATTAGTTGTAAGATGGAGTCAATTTGAGTCTCTTAATAGCCATGATCATATACCA 406
Qy 3659 agtgaagctggtggaacatatggtctccattttacagttaagggaataataatggagcagatt 3718
Db 405 AGTGAAGCTGTTGGAACATATGGTCTCCATTTTACAGTTAAGGAATATAATGGACAGATT 346
Qy 3719 aatatgttctgtctcagcccaaatcccttcttaaggaagactgccctactatagcag 3778
Db 345 AATATTGTTCTGTCTGATGCCCAATCCCTTTCTAAGGAAGACTGCCCTACTATAGCAG 286
Qy 3779 ttttattattgtcaatttatgaataataatgaatgagagttcttggtacccctctgtcttt 3838
Db 285 TTTTATATTGTCAATTTATGAATAATAATGAATGA-GAGTTCTGTTACCTCCTGCTCTT 227
Qy 3839 acaaatattgggtgtgtgtccagtatatttcccttttaaacomttcccaattccgggtgtgt 3898
Db 226 ACAATATTGG-----TGTGTGTCAGTATTTTTCCTTTTAAACCATCCAATCGGTGTGTAG 171
Qy 3899 aggtggatgtttccatttgggttttaatttgtatatccctgatactagctattgggtcat 3958
Db 170 TGAG-----TTTCATTTTGGTTTTTAATTGTATATCCCTGATAGCTATAATTTGGTCAAT 116
Qy 3959 agaatcttttatcacattctagatgcagctctcttgcggatatacactgattgagatatta 4018
Db 115 AGAAATCTTTATACATTCATAGATGCAATCTCTCTGTCGGATATATGTATTGAGATATTA 56
Qy 4019 cacctagtgtgggtgactgttttcttttatgtcttttggatgaatgaagaatttt 4073
Db 55 CACCTAGTCTGTGGCTTGACTGTTTTTCTTTTATGTCTTTTGTGATGAATAGAAGTTTT 1

RESULT 14
AX341490
LOCUS AX341490 544 bp DNA linear PAT 10-JAN-2002
DEFINITION Sequence 1737 from Patent WO0196388.
ACCESSION AX341490
VERSION AX341490.1 GI:18137472
KEYWORDS human.
SOURCE


```
/evidence-experimental
3289. .3423
/rpt_family="L1MC/D"
complement(join(4839. .5019,6434. .6483,12661. .12774,
15548. .15703,17232. .17300,22652. .23099,24691. .24963,
28133. .28262,33818. .33915,39772. .39965,40581. .40698,
47628. .47731,50142. .50301,50783. .50931,52130. .52263,
53933. .54020))
/gene="NBS1"
/notes="NBS1"
/notes="NBS1, 2.4kb transcript"
/evidence-experimental
complement(4854. .4859)
/gene="NBS1"
/notes="NBS1, Nibrin, 2.4kb transcript"
/evidence-experimental
complement(join(4989. .5019,6434. .6483,12661. .12774,
15548. .15703,17232. .17300,22652. .23099,24691. .24963,
28133. .28262,33818. .33915,39772. .39965,40581. .40698,
47628. .47731,50142. .50301,50783. .50931,52130. .52263,
53933. .53969))
/gene="NBS1"
/codon_start=1
/evidence-not_experimental
/product="Nibrin"
/protein_id="AAD08722.1"
/db_xref="GI:3108229"
/translation="MKMLPAAGPAGGEPYRLLTGYEVVGVKKNCAILLIENDQSIERN
HAYLTANFVNLSDDEIPVLTKDNSKYGTFTVNEEKWONGFSRTLKSGDGTFGVF
GSKFRTEYPLVACSSLDVSGKTALNQATLOLGGTFTVNNWTECHLMVSVKVIK
TICALICGRPIVKPEYFTEFLKAVESKKOPPOIESYPLDERSIGSKNDLSGRBR
KQIFKGTIFILNAKHKLSSAVFVGGEARLITTEENEEHNFLLAPGTCVVDTGIT
NSQTLIDPCQKWIQSIDMLQROGLRPIPEADTGLAVIPMTTKNYCDPGHPSGLK
TTTTGSLQGVSVDEKILMSPAPVNTTYYVADTSEADTDLSEPKETKVKMSQOK
FRLMSQDAPTVKESCTSSNNNSMWNTLAKMRIPNYQSLPTKIPSTINKSKDRASQOK
OTNSIRNYFQPTNFKRDERDEENQESSKSARLETSCSLEEQPATPSLWKKQOHL
SENEPVDTSNNLFTDLDLSIVKNSAKSHAAKLSRKNKREMDOVALEDVLEQL
FKDKTELEIDVYKQOEEDVNVKRRPRMDIETNDFSDEAVPESSKISOENEDYGR
ELKEDSLWSAKEISNKLQDDSEMLPKLLLTFRSLVLIKNTSRNPSSGINDDYGL
KNPKFKKVTYTPGAGKLPHIIGGSDLIHARKNTELEWLRQEMEVEVQNHAKKEESLA
DDLFRYNPLYKRRR"
complement(5495. .5758)
/rpt_family="AluSg"
7018. .7204
/rpt_family="L2"
complement(7385. .7435)
/rpt_family="L2"
7620. .7799
/rpt_family="MER6B"
complement(7891. .8070)
/rpt_family="MIR"
complement(9507. .9828)
/rpt_family="AluJb"
complement(10262. .10401)
/rpt_family="MIR"
22006. .12320
/rpt_family="L1MEC"
complement(13497. .13554)
/rpt_family="AluJb"
13555. .14039
/rpt_family="MER84"
complement(14040. .14287)
/rpt_family="AluJb"
complement(16605. .16705)
/rpt_family="MIR"
16756. .17038
/rpt_family="AluSp"
17039. .17065
/rpt_family="(TAAA)n"
17357. .17391
/rpt_family="U2"
18286. .18461
/rpt_family="L2"
18577. .18720
```

```
repeat_region
19368. .19467
/rpt_family="L1MD1"
19606. .19938
/rpt_family="MER2"
20473. .20739
/rpt_family="AluY"
21689. .21772
/rpt_family="MER5A"
complement(21710. .21800)
/rpt_family="MER5A"
complement(24179. .24388)
/rpt_family="L2"
26114. .26192
/rpt_family="MER58A"
26204. .27207
/rpt_family="Tiger3b"
complement(27208. .27713)
/rpt_family="MER1A"
27617. .104987
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clonene="284N21"
27714. .27838
/rpt_family="Tiger3b"
27863. .27988
/rpt_family="MER58A"
28879. .29181
/rpt_family="AluJb"
29184. .29341
/rpt_family="MIR"
29364. .29487
/rpt_family="L2"
29780. .30050
/rpt_family="AluJo"
30058. .30077
/rpt_family="(TAA)n"
30123. .30245
/rpt_family="L2"
32351. .32517
/rpt_family="MIR"
33083. .33362
/rpt_family="L2"
33351. .33683
/rpt_family="L2"
```

Query Match 10.2%; Score 451.2; DB 9; Length 330250;
Best Local Similarity 94.1%; Pred. No. 2.2e-71;
Matches 2024; Conservative 9; Mismatches 26; Indels 91; Gaps 6;

```
Qy 2278 tcttttagatacaatccttatttataaaggagaagataactgaggattttataaagaag 2337
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 5028 TCITTCAGATACACTCCTTATTATAAAGGAGAGATACTGAGGATTTAAAGAAG 4969
Qy 2338 ccatggaaaaacttcctagtaagcatctacttcaggccaacaaggttatatgaatatata 2397
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4968 CCATGGAAAAAATTCCTAGTAGCATCTACTTCAGGCCAACCAAGGTATATGAATATA 4909
Qy 2398 gtgtatagaagcatttaagttacaatgtttttatggcctaataattataaataatgca 2457
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4908 GTGTATAGAAGCGATTTAAGTTACAATGTTTTATGGGCTAAATTTATTAATAAATGCA 4849
Qy 2458 caaaactttgattctttttagtatgtaacaattgtttgtctgttttcagggtttgtcattg 2517
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4848 CAAACTTTGATCTTTTGTATGTATGTAACAATTTGTTTGTCTGTTTCAGGCTTTGTCATTG 4789
Qy 2518 catcttttttcatcttttaaatgtgtttttgttttataaataagttatagtcacagttc 2577
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4788 CATCTTTTTCATTTTAAATGTGTTTGTGTTTATTAATAGTTAAATAGTACAGTTC 4729
Qy 2578 aaaaattcctaattacgaaggttaaggaactaaagtcaccctccaccattgtcctagct 2637
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4728 AAAATTCTAAATGTACGTAAGGTAAA-GACTAAAGTCACCCCTCCACCTGCTAGCT 4670
```

QY 2638 act----- 2640
 Db 4669 ACTTGGTTCCCTCAGAAAAAATTCATGATCTCATCTTCTTATGAATCTTTCCAGGGAT 4610
 QY 2641 -----tattttaaataattctctacacaaatgatagataac 2678
 Db 4609 TTTTGTAGTCTCTATTCAAATTCCTATTTTTTAAATAATTTCTTACACAAATGATGATTAAC 4550
 QY 2679 atatgaagtgcttctacacccctgtcttttacttagtaagaattaaaaattatagaatc 2738
 Db 4549 ATATGAGTGTCTACACCTTGCTTTTTTACTTAGT-AGATTTAAAAATTTATAGGAATATC 4491
 QY 2739 aatataatgttttaataattttcttcttccattatctgtagctcttaccataaacttgg 2798
 Db 4490 AATATAATGTTTTTTAATATTTTTTCTTTTCCATTTATGCTGTAGTCTTACCTAAACTCTGG 4431
 QY 2799 tgatccaaacaaatggcttcagtggtgcaagatgctcaactacatgttatcttagtaac 2858
 Db 4430 TGATCCAAACAAATGGCTTCAGTGGTGCAGATGTCACCTACATGTTTATCTAGTACTAG 4371
 QY 2859 aaactgaagaccatgtgagactctcatcaacatgggttttagttttcaccagaatggaaa 2918
 Db 4370 AAATGAAGACCATGTGAGACTTCATCAACATGGGTTAGTTTTCCACGAAATGGAAA 4311
 QY 2919 gaoctgtaccccttttgggtgttactgagctgggtgggtgtctgttttgagcttatt 2978
 Db 4310 GACCTGTACCCCTTTTTTGGTGGTCTTACTGAGCTGGGTGGTGTCTGTTTGGAGCTTATT 4251
 QY 2979 tagagtcctagtgttctacttataaagtagaagatggatgtgtgtgttttcttcttacc 3038
 Db 4250 TAGAGTCTAGTTTCTACTTATAAAGTAGAATGTGAGATGTTTCTTTTCTTACC 4191
 QY 3039 ktaaaggagatggtgaagaacaaatgaatgtcttttttcaaacctttattgacaagtatt 3098
 Db 4190 TTAAGGGAGATGATAGAACAAATGAATGCTTTTTTCAACCTTATTGACAACTGATT 4131
 QY 3099 ttcaagtcgtgttccaaaaatatattcatgtacgtgtgtatccagcaagaaggagttcca 3158
 Db 4130 TTCAAGTCTGTGTTTCAAAAAATATATTCTGTACCTGTGATCCAGCAAGAGGGAGTTCCA 4071
 QY 3159 gtcaagagtcactacactgattagttgttttagagaatgagaatggaacagtgaggat 3218
 Db 4070 GTCAAGGTCATCACAACACTGATTAGTTGTTTGTAGAGAATGAGAAATGGAACAGTGAAGAT 4011
 QY 3219 ggaggccatatttccatgacttccctgttaacagaagcaacagaagggagagctg 3278
 Db 4010 GGAGGCCATATTTCCATGACTTCCCTTGTAAACAGAGCAACAGAGGGAGGCTG 3951
 QY 3279 gcctctacatcactctcaccttccaaatcttgttgaagtgcatctacttgcagaaaccaa 3338
 Db 3950 GCCTCTACATCACTCTCACCTTCCAAATCTTGTGGAAGTGCATCTACTTGCAGAACCAA 3891
 QY 3339 attaaacttaacttccaaagtctgctgtcagggtgaaactccagctgcaaggagtag 3398
 Db 3890 ATTAACCTTACTTCCAAAGTCTGGCTGTGAGGTGGAATCCAGCTGCAAGGGAGTAG 3831
 QY 3399 ggaatgaaggtcttttttaaagcttctcagccttccagtagggaacagaataatgggtga 3458
 Db 3830 GGAATGAAGGTCTTTTTTAAAGCTTCTCAGCCTTCTTAGGGAAACAGAAATTTGGGTGA 3771
 QY 3459 gccaatctgcaattctactacagggcattgagaccagttagattattgaaatattataga 3518
 Db 3770 GCCAATCTGCAATTTCTTACTACAGGCATTTGAGACCAGTTAGATTATTGAAATATTATAGA 3711
 QY 3519 gagttaagaacttaattatgatagtggtatgacatttgatagaacatgggatacttt 3578
 Db 3710 GAGTTATGAACACTTAAATTTATGATAGTGGTATGACATTTGATAGAACATGGGATACTTT 3651
 QY 3579 agaagtagaattgacagggcatatttagttgatgaaatggagtcatttgagctctyatta 3638
 Db 3650 AGAAGTAGAATTGACAGGGCATATTAGTTGATGAAATGGAGTCATTTGAGTCTCTTAATA 3591

QY 3639 gccatgtatcataattaccgaagtgaagctggtggaacatatggtctccatttttacagtta 3698
 Db 3590 GCCATGTATCATATAATTACCAAGTGAAGCTGGTGAACATATGGTCTCCATTTTACAGTTA 3531
 QY 3699 aggaataataatggacagatataattgttctctcatcgtccacaaatcccttttctaagga 3758
 Db 3530 AGGAATAATAATGGACAGATAAATTTGTCTCTCATGATGCCCAAAATCCCTTTTAAAGGA 3471
 QY 3759 agactgcctactactatagcagtttttatttattgttcaatttatgaataataatgaatgagag 3818
 Db 3470 AGACTGCCCTACTATAGCAGTTTTTATATTGTCAATTTATGAATATATGAATGA-GAG 3412
 QY 3819 ttctgttacctctctgtctttttacaaatattgggtgtgtgtccagttatttttcccttttaac 3878
 Db 3411 TTCTGGTACCTCCCTGCTCTTTACAAATATTGG----TGTGTGCAGTATTTTTCCTTTTAA 3356
 QY 3879 cmttcccaattcgggtgtgtgtaggtttccatttgggttttcaatttgaatttatccct 3938
 Db 3355 CCATTCCAATCGGTGTGTAGTGATG----TTTCATTTTGGTTTTTAATTTGTATATCCCT 3301
 QY 3939 gatagctataattgggtcatagaaattctttatcacattcttagatgcaagtcctcttgcgg 3998
 Db 3300 GATAGCTATAAATGGGTCAATAGAAATCTTTTATACATTTCTAGATGCAAGTCTCTTGC 3241
 QY 3999 atatacgtattgagatattacacacttagctctgtggtgactgttttcttttgaictttg 4058
 Db 3240 ATATATGATTGAGATATTACACCTAGTCTGTGGCTTGTGACTGTTTTCTTTATGCTTTTG 3181
 QY 4059 atgaatagaagttttaaatttttgacaaggtcaaaattttttttttttttttttttgata 4118
 Db 3180 ATGAATAGAAGTTTTAAATTTTGACAAGGTCAAAATTTATTTTCTTTTGTGTGATATT 3121
 QY 4119 tttctctccaatttaaccccaagattttcagatatctctctctattatataaaactttata 4178
 Db 3120 TTTTCTCTCCAATTTAACCCCAAGATTTTCAGATATTCTGCTCTATTATATAAACTTTATA 3061
 QY 4179 tttttattttgtgatctacacttgaattgatgatgatgtgtgtgaattatgatcagggctt 4238
 Db 3060 TTTTATATTTGTGATCTACCTTGAATTTGATGATGATGTTGTGATTTATGATCAGGGTT 3001
 QY 4239 cttttttcccccatacaagtatccagtcattgtaaacactgtttattgaaagaattatcc 4298
 Db 3000 CTTTTTTTCCCCCAYACAAGTATCCAGTCATTTGAACACTGTTTATTTGAAAGAATATATCC 2941
 QY 4299 tttctctattaaattacccttgcatttagtaaaaaatcaatttaaccatrm 4348
 Db 2940 TTTCTCTCATTAATAATTACCTTGCCTCAATTAGTAAAAAATCAATTAACCATAA 2891

Search completed: August 15, 2002, 00:14:39
 Job time: 10004 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 01:40:40 ; Search time 66.52 Seconds
(without alignments)
1960.887 Million cell updates/sec

Title: US-09-837-602-2

Perfect score: 3899

Sequence: 1 MWKLLPAAGPAGGEPYLLF.....KEESLADDLFRYPYLKRRR 754

Scoring table:

BLOSUM62

Gapop: 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_l9.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|--------|---------------|--------|-----------|--------------------|
| 1 | 3896 | 99.9 | 754 | 4 O60934 | O60934 homo sapien |
| 2 | 2710 | 69.5 | 751 | 11 Q9R207 | Q9R207 mus musculu |
| 3 | 2702 | 69.4 | 751 | 11 Q88981 | Q88981 mus musculu |
| 4 | 2705 | 69.3 | 751 | 11 Q9R1X1 | Q9R1X1 mus musculu |
| 5 | 2685.5 | 68.9 | 750 | 11 Q9JIL9 | Q9JIL9 rattus norv |
| 6 | 1766.5 | 45.3 | 753 | 13 Q9DE07 | Q9DE07 gallus gall |
| 7 | 473 | 12.1 | 811 | 5 Q9VT40 | Q9VT40 drosophila |
| 8 | 227.5 | 5.8 | 1460 | 4 Q75058 | Q75058 homo sapien |
| 9 | 227.5 | 5.8 | 1486 | 4 Q9UQ08 | Q9UQ08 homo sapien |
| 10 | 215.5 | 5.5 | 1510 | 5 Q25920 | Q25920 plasmodium |
| 11 | 210.5 | 5.4 | 1584 | 4 Q9UQ09 | Q9UQ09 homo sapien |
| 12 | 195.5 | 5.0 | 1661 | 5 Q06166 | Q06166 plasmodium |
| 13 | 187.5 | 4.8 | 2083 | 5 Q9N435 | Q9N435 caenorhabdi |
| 14 | 185.5 | 4.8 | 3147 | 5 Q17464 | Q17464 caenorhabdi |
| 15 | 185.5 | 4.8 | 3484 | 5 P91257 | P91257 caenorhabdi |
| 16 | 181.5 | 4.7 | 2081 | 10 Q9LH98 | Q9LH98 arabidopsis |

| | | | | | | |
|----|-------|-----|-------|----|--------|--------------------|
| 17 | 179.5 | 4.6 | 1790 | 3 | Q07380 | Q07380 saccharomyc |
| 18 | 177 | 4.5 | 1211 | 5 | Q23260 | Q23260 caenorhabdi |
| 19 | 176.5 | 4.5 | 1265 | 4 | Q9P274 | Q9P274 homo sapien |
| 20 | 176.5 | 4.5 | 1323 | 5 | Q9NB35 | Q9NB35 plasmodium |
| 21 | 175 | 4.5 | 823 | 4 | Q9H0H6 | Q9H0H6 homo sapien |
| 22 | 174 | 4.5 | 725 | 5 | Q965E8 | Q965E8 entamoeba h |
| 23 | 174 | 4.5 | 911 | 3 | Q06704 | Q06704 saccharomyc |
| 24 | 174 | 4.5 | 2603 | 6 | Q9XS53 | Q9XS53 potorous tr |
| 25 | 174 | 4.5 | 17352 | 5 | Q95YM2 | Q95YM2 procamburus |
| 26 | 172.5 | 4.4 | 1038 | 13 | Q90784 | Q90784 gallus gall |
| 27 | 171 | 4.4 | 1812 | 5 | Q9VZD9 | Q9VZD9 drosophila |
| 28 | 171 | 4.4 | 3257 | 5 | Q9V736 | Q9V736 drosophila |
| 29 | 169.5 | 4.3 | 2287 | 5 | Q19761 | Q19761 caenorhabdi |
| 30 | 169 | 4.3 | 1229 | 5 | Q94185 | Q94185 caenorhabdi |
| 31 | 169 | 4.3 | 1616 | 4 | Q96PH3 | Q96PH3 homo sapien |
| 32 | 168.5 | 4.3 | 1125 | 16 | Q98PM9 | Q98PM9 mycoplasma |
| 33 | 168.5 | 4.3 | 3704 | 5 | P91904 | P91904 caenorhabdi |
| 34 | 168 | 4.3 | 594 | 3 | Q9P4A3 | Q9P4A3 emericella |
| 35 | 168 | 4.3 | 808 | 4 | O15083 | O15083 homo sapien |
| 36 | 168 | 4.3 | 832 | 4 | Q9NUJ4 | Q9NUJ4 homo sapien |
| 37 | 168 | 4.3 | 1185 | 4 | Q9NQL1 | Q9NQL1 homo sapien |
| 38 | 168 | 4.3 | 1235 | 4 | Q9H2G2 | Q9H2G2 homo sapien |
| 39 | 167.5 | 4.3 | 456 | 11 | Q9CS77 | Q9CS77 mus musculu |
| 40 | 167.5 | 4.3 | 853 | 10 | Q9LFE4 | Q9LFE4 arabidopsis |
| 41 | 167.5 | 4.3 | 1287 | 5 | Q9NGS5 | Q9NGS5 dictyosteli |
| 42 | 167.5 | 4.3 | 2363 | 4 | Q99968 | Q99968 homo sapien |
| 43 | 167 | 4.3 | 1327 | 4 | Q9Y2L2 | Q9Y2L2 homo sapien |
| 44 | 167 | 4.3 | 1395 | 2 | Q9AIS0 | Q9AIS0 staphylococ |
| 45 | 166.5 | 4.3 | 927 | 3 | Q9HES9 | Q9HES9 emericella |

ALIGNMENTS

RESULT 1

| | | | |
|--------|--|---|---------|
| O60934 | PRELIMINARY; | PRT; | 754 AA. |
| ID | O60934 | O60672; | |
| AC | O60934 | (TREMBLrel. 07, Created) | |
| DT | 01-AUG-1998 | (TREMBLrel. 07, Last sequence update) | |
| DT | 01-AUG-1998 | (TREMBLrel. 07, Last sequence update) | |
| DE | 01-JUN-2001 | (TREMBLrel. 17, Last annotation update) | |
| DE | NIBRIN (NIJMEGEN BREAKAGE SYNDROME PROTEIN 1) (CELL CYCLE REGULATORY | | |
| DE | PROTEIN P95). | | |
| GN | NBS1 OR NBS. | | |
| OS | Homo sapiens (Human). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| OX | NCBI_TaxID=9606; | | |
| RN | [1] | | |
| RX | SEQUENCE FROM N.A., AND VARIANT GLN-553. | | |
| RX | MEDLINE=98250062; PubMed=9590180; | | |
| RA | Varon R., Vissinga C., Platzer M., Cerosaletti K.M., Chrzanoska K.H., | | |
| RA | Saar K., Beckmann G., Seemanova E., Cooper P.R., Nowak N.J., Stumm M., | | |
| RA | Weenae C.M.R., Gatti R.A., Wilson R.K., Digweed M., Rosenthal A., | | |
| RA | Sperling K., Concannon P., Reis A. | | |
| RT | "Nibrin, a novel DNA double-strand break repair protein, is mutated in | | |
| RT | Nijmegen breakage syndrome."; | | |
| RL | Cell 93:467-476(1998). | | |
| RN | [2] | | |
| RN | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=98282099; PubMed=9620777; | | |
| RA | Matsuura S., Tauchi H., Nakamura A., Kondo N., Sakamoto S., Endo S., | | |
| RA | Smeets D., Solder B., Belohradsky B.H., Kaloustian V.M., Oshimura M., | | |
| RA | Isonura M., Nakamura Y., Komatsu K., | | |
| RT | "Positional cloning of the gene for Nijmegen breakage syndrome."; | | |
| RT | Nat. Genet. 19:179-181(1998). | | |
| RN | [3] | | |
| RN | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=99134304; PubMed=9933573; | | |
| RA | Tauchi H., Matsuura S., Isonura M., Kinjo T., Nakamura A., | | |
| RA | Sakamoto S., Kondo N., Endo S., Komatsu K., Nakamura Y.; | | |
| RT | "Sequence analysis of an 800-kb genomic DNA region on chromosome 8q21 | | |
| RT | that contains the Nijmegen breakage syndrome gene, NBS1."; | | |

Genomics 55:242-247(1999).
 [4]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=98250063; PubMed=9590181;
 RA Carney J.P., Maser R.S., Olivares H., Davis E.M., Le Beau M.,
 Yates J.R. III, Hays L., Morgan W.F., Petrini J.H.J.;
 RT "The hMre11/hRad50 protein complex and Nijmegen breakage syndrome:
 RT linkage of double-strand break repair to the cellular DNA damage
 response.";
 RL Cell 93:477-486(1998).
 CC -1- FUNCTION: INVOLVED IN REPAIR OF DNA DOUBLE-STRAND BREAKS. WORKS
 CC AS PART OF A COMPLEX WHICH ALSO INCLUDES RAD50 AND MRE11.
 CC -1- TISSUE SPECIFICITY: UBQUITOUS. EXPRESSED AT HIGH LEVELS IN
 CC TESTIS.
 CC -1- DISEASE: DEFECTS IN NBS1 ARE THE CAUSE OF NIJMEGEN BREAKAGE
 CC SYNDROME, AN AUTOSOMAL RECESSIVE CHROMOSOMAL INSTABILITY SYNDROME
 CC CHARACTERIZED BY MICROCEPHALY, GROWTH RETARDATION,
 CC IMMUNODEFICIENCY, CANCER PREDISPOSITION, CELL CYCLE CHECKPOINT
 CC DEFECTS AND IONIZING RADIATION SENSITIVITY.
 DR EMBL: AF051334; AAC39732.1; -;
 DR EMBL: AB013139; BAA28616.1; -;
 DR EMBL: AF049895; AAD08722.1; -;
 DR EMBL: AF069291; AAC62232.1; -;
 DR EMBL: AF058696; AAC39752.1; -;
 DR MIM: 251260; -;
 DR MIM: 602667; -;
 DR InterPro: IPR001357; BRCT.
 DR InterPro: IPR000253; FHA_domain.
 DR Pfam: PF00533; BRCT; 1.
 DR Pfam: PF00498; FHA; 1.
 DR SMART: SM00292; BRCT; 1.
 DR SMART: SM00240; FHA; 1.
 DR PROSITE: PS50006; FHA_DOMAIN; 1.
 KW DNA repair; Glycoprotein; Polymorphism.
 FT VARIANT 185 185 E -> Q.
 FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 142 142 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 440 440 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 600 600 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 666 666 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 754 AA; 84958 MW; CD602F09BA73DAB6 CRC64;
 Query Match 99.9%; Score 3896; DB 4; Length 754;
 Best Local Similarity 99.9%; Pred. No. 1.7e-220;
 Matches 753; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MWKLLPAAGPAGGEPYRLLTGVEYVVGKRNKCAILLIENDQSIIRNHAVALTANFVSNLSQT 60
 DB 1 MWKLLPAAGPAGGEPYRLLTGVEYVVGKRNKCAILLIENDQSIIRNHAVALTANFVSNLSQT 60
 QY 61 DEIPVLTKDNSKYGFVNEEKQNGFSRTLKSGDGTITFGVFGSKFRIEYPLVACSSCL 120
 DB 61 DEIPVLTKDNSKYGFVNEEKQNGFSRTLKSGDGTITFGVFGSKFRIEYPLVACSSCL 120
 QY 121 DVSGKTALNAQIILQGGFTVNNTEECTHLVMVSVKVTIKTICALICGRPIVAPYFTF 180
 DB 121 DVSGKTALNAQIILQGGFTVNNTEECTHLVMVSVKVTIKTICALICGRPIVAPYFTF 180
 QY 181 LKAVQSKKQPPIESFYPLDEPISGSKNVDSLGRQERKQIFKGTIFILNAKQHKLLS 240
 DB 181 LKAVQSKKQPPIESFYPLDEPISGSKNVDSLGRQERKQIFKGTIFILNAKQHKLLS 240
 QY 241 AVVFGGEARLITEENEEHNFLLAPGTCVVDGTGINTNSQTLIPDCQKKWTQSDMLQ 300
 DB 241 AVVFGGEARLITEENEEHNFLLAPGTCVVDGTGINTNSQTLIPDCQKKWTQSDMLQ 300
 QY 301 GLRPIPEAEIGLAVIFMTTKNYCDPOGHPSTGLKTTTPGSLSQGVSDKLMPSAPVNT 360
 DB 301 GLRPIPEAEIGLAVIFMTTKNYCDPOGHPSTGLKTTTPGSLSQGVSDKLMPSAPVNT 360

QY 361 TTYVADTESQADTWLDSERPKEIKVSKMBQKFRMLSQDAPTVEKESCKTSSNNNSMYSNT 420
 DB 361 TTYVADTESQADTWLDSERPKEIKVSKMBQKFRMLSQDAPTVEKESCKTSSNNNSMYSNT 420
 QY 421 LAKMRIPNYQLSPTKLPSINKSKDRASQOQQTNSIRNYFOPSTKKRERDEENQEMSSCKS 480
 DB 421 LAKMRIPNYQLSPTKLPSINKSKDRASQOQQTNSIRNYFOPSTKKRERDEENQEMSSCKS 480
 QY 481 ARLETSCSLLEQTPATPSPSLWKNKEQHLSENEPVDNTSDNNLFTDTDLKSTIVKNSASKSH 540
 DB 481 ARLETSCSLLEQTPATPSPSLWKNKEQHLSENEPVDNTSDNNLFTDTDLKSTIVKNSASKSH 540
 QY 541 AAEKLSRNKKRMDVAIEDEVLEQLFKDTKPELEIDVKVQKQEDYVNRKRPRMDIETN 600
 DB 541 AAEKLSRNKKRMDVAIEDEVLEQLFKDTKPELEIDVKVQKQEDYVNRKRPRMDIETN 600
 QY 601 DTFSDEAVPSSKISQENEIGKRELKEDSLWSAKEISNNNDKLQDDSEMLPKLLLTFR 660
 DB 601 DTFSDEAVPSSKISQENEIGKRELKEDSLWSAKEISNNNDKLQDDSEMLPKLLLTFR 660
 QY 661 SLVIKNSTSRNPGINDDYQLKFNFKFKVTPYAGKLPHIIGGSDLIHAHARKNTELE 720
 DB 661 SLVIKNSTSRNPGINDDYQLKFNFKFKVTPYAGKLPHIIGGSDLIHAHARKNTELE 720
 QY 721 EWLROQMEVQNHAKESLADDLFRYPNPKRRR 754
 DB 721 EWLROQMEVQNHAKESLADDLFRYPNPKRRR 754
 RESULT 2
 Q9R207 PRELIMINARY; PRT; 751 AA.
 ID Q9R207
 AC Q9R207
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE NIBRIN.
 GN NBN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20108791; PubMed=10640816;
 RA Vissinga C.S., Yeo T.C., Woessner J., Massa H.F., Wilson R.K.,
 RA Trask B.J., Concannon P.;
 RT "Identification, characterization, and mapping of a mouse homolog of
 RL the gene mutated in Nijmegen breakage syndrome.";
 RL Cytogenet. Cell Genet. 87:80-84(1999).
 DR EMBL: AF076687; AAD20943.1; -;
 DR MGD; MGI:1351625; Nbn.
 DR InterPro: IPR001357; BRCT.
 DR InterPro: IPR000253; FHA_domain.
 DR Pfam: PF00533; BRCT; 1.
 DR Pfam: PF00498; FHA; 1.
 DR SMART: SM00292; BRCT; 1.
 DR SMART: SM00240; FHA; 1.
 DR PROSITE: PS50006; FHA_DOMAIN; 1.
 SQ SEQUENCE 751 AA; 83794 MW; C9F597CC08227B2C CRC64;
 Query Match 69.5%; Score 2710; DB 11; Length 751;
 Best Local Similarity 70.9%; Pred. No. 5.7e-151;
 Matches 537; Conservative 81; Mismatches 129; Indels 10; Gaps 9;
 QY 1 MWKLLPAAGPAGGEPYRLLTGVEYVVGKRNKCAILLIENDQSIIRNHAVALTANFVSNLSQT 60
 DB 1 MWKLLPAAGPAGGEPYRLLTGVEYVVGKRNKCAILLIENDQSIIRNHAVALTANFVSNLSQT 60
 QY 61 DEIPVLTKDNSKYGFVNEEKQNGFSRTLKSGDGTITFGVFGSKFRIEYPLVACSSCL 120
 DB 61 DEIPVLTKDNSKYGFVNEEKQNGFSRTLKSGDGTITFGVFGSKFRIEYPLVACSSCL 120

Db 61 DEIPTLTIKDNSKYGTFFVNEEKMQTGLSCTLKTDGRTVFGVFSKFRVEYEPVLVGCSSCL 120
QY 121 DVSGKTALNOAILQLGGFTVNNWTECTHLVMSVKTITKICALICGRPIVKPEYFTEF 180
Db 121 DVSGKTALNOAILQLGGFTVNNWTECTHLVMSVKTITKICALICGRPIVKPEYFSEF 180
QY 181 LKAVOSKKOPPOIESYPPPLDEPISGSKNVDSGROKQIFKGTFFILNAKOHKKLSS 240
Db 181 LKAVESKKOPPDIESYPPIDEPAIGSKSVDSLSGRHERKQIFKGTFFVFLNAKOHKKLSS 240
QY 241 AVVFGGEARLITEENEEHNFFLAGPCTCVDGTGITSOTLIPDCQKWTQSTMDMLQRO 300
Db 241 AVVFGGEARLMAEDDEEESQFSAPGTCVVDVGTGITSOTLIPDCQKWTQSTMDMLQRO 300
QY 301 GLRPIPEAIGLAVIFMTTKNYCDPOGHPSTGLKTTTPGPSLQGVSDVKLMPSPVNT 360
Db 301 GLRPIPEAIGLAVIFMTTENYCNPOGQCTELKTTTPGPSLQGVLSANGKIIPSPVNM 360
QY 361 TTYVADTESPADT-WDLSEPKKEIVSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVSN 419
Db 361 TTYVADTESPADT-CPLSERPEEVKIPGLEQSSRKLSEOTFNKEAPKPSKANNVAD 420
QY 420 TLAKMRIPNYQSLPTKLPISNKSKRASQOQNTSNRTNFQPSKTKRDERDEENQMSCK 479
Db 421 TLVRGKTPSYQLSPMKFPVANKNKDWTSSQOO-NSIKNYFQPCTRKREDDEDNPELSSCK 479
QY 480 SARIETSCSLEQTOPATPSLWKNKEQHLSENPVDNNDNLTDTDLKSIYKNSASKS 539
Db 480 SSRMELSCSLEQTOPAGPSLWKSKE-HQSONATLDREADTSSVGGMDIELNRKSPDRKP 538
QY 540 HAAEKLRSNKKREMDVVAIEDEVLEQLFKDTKPELEIDVVKQOEEDVNVKRPMDIET 599
Db 539 LPTETLPRKRKDV-DLATEEVEVLEELLRSTKPELAVQVKEQEAADTIRKKPRMDAER 597
QY 600 NDTFSDAEVPESSKISQENIEGKRELKEDSLWSAK-EISNNDKLODDSEMLPKLLLTE 658
Db 598 NRPLNGGSEPESSALQEDEREKKDELQTES-WSTRKHEIANSGLQDSEELPRKLLTE 656
QY 659 FRSLVTK--NSTSRNPSGINDGYQLKNFKKFKKVTYPGAGKLPHIIGSDLIHAHARKN 716
Db 657 FRSLVSNHNSTSRNLIC-VN-ECGPLKNFKFKKATFPGAGKLPHIIGSDLVGHARKN 714
QY 717 TELEEWLRQEMEVQONHAKESLADDLFRYNPYLKR 753
Db 715 TELEEWLKQEMEVQKQAKESLADDLFRYNPNVKKR 751

RESULT 3
ID O88981 PRELIMINARY; PRT; 751 AA.
AC O88981;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE NIBRIN.
GN NBN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10990;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN STEM;
RA Mas C., Bourgeois F., Simonneau M.;
RT "Isolation of 50 cDNAs differentially expressed in embryonic forebrain
as compared to mid and hindbrain : a strategy to identify candidate
genes involved in human neurodevelopmental diseases."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF092840; AAC62113.1;
DR MGD; MGI:1351625; Nbn.
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR000253; FHA_domain.
DR Pfam; PF00533; BRCT; 1.

DR Pfam; PF00498; FHA; 1.
DR SMART; SM00292; BRCT; 1.
DR SMART; SM00240; FHA; 1.
DR PROSITE; PS50006; FHA_DOMAIN; 1.
SQ SEQUENCE 751 AA; 83826 MW; 9D136BBC23DC51D9 CRC64;

Query Match 69.4%; Score 2705; DB 11; Length 751;
Best Local Similarity 70.8%; Pred. No. 11e-150;
Matches 536; Conservative 82; Mismatches 129; Indels 10; Gaps 9;

QY 1 MWKLLPAAGPAGGEPYRLLTGVVEYVVGKKNCAILINDQSISRNHAVLTANFVSNLSQT 60
Db 1 MWKLLPAAGAGSLGEPYRLLAGVEYVVGKKNCGILINDQSISRNHAVLTANFVSNLSQT 60
QY 61 DELPVLTLKDNSKYGTFFVNEEKMGNGFSRTLKSGDGTTFGVGSKFRIEYEPVLVACSCL 120
Db 61 DEPTLTIKDNSKYGTFFVNEEKMQTGLSCTLKTDGRTVFGVFSKFRVEYEPVLVGCSSCL 120
QY 121 DVSGKTALNOAILQLGGFTVNNWTECTHLVMSVKTITKICALICGRPIVKPEYFTEF 180
Db 121 DVSGKTALNOAILQLGGFTVNNWTECTHLVMSVKTITKICALICGRPIVKPEYSEF 180
QY 181 LKAVOSKKOPPOIESYPPPLDEPISGSKNVDSGROKQIFKGTFFILNAKOHKKLSS 240
Db 181 LKAVESKKOPPDIESYPPIDEPAIGSKSVDSLSGRHERKQIFKGTFFVFLNAKOHKKLSS 240
QY 241 AVVFGGEARLITEENEEHNFFLAGPCTCVDGTGITSOTLIPDCQKWTQSTMDMLQRO 300
Db 241 AVVFGGEARLMAEDDEEESQFSAPGTCVVDVGTGITSOTLIPDCQKWTQSTMDMLQRO 300
QY 301 GLRPIPEAIGLAVIFMTTKNYCDPOGHPSTGLKTTTPGPSLQGVSDVKLMPSPVNT 360
Db 301 GLRPIPEAIGLAVIFMTTENYCNPOGQCTELKTTTPGPSLQGVLSANGKIIPSPVNM 360
QY 361 TTYVADTESPADT-WDLSEPKKEIVSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVSN 419
Db 361 TTYVADTESPADT-CPLSERPEEVKIPGLEQSSRKLSEOTFNKEAPKPSKANNVAD 420
QY 420 TLAKMRIPNYQSLPTKLPISNKSKRASQOQNTSNRTNFQPSKTKRDERDEENQMSCK 479
Db 421 TLVRGKTPSYQLSPMKFPVANKNKDWTSSQOO-NSIKNYFQPCTRKREDDEDNPELSSCK 479
QY 480 SARIETSCSLEQTOPATPSLWKNKEQHLSENPVDNNDNLTDTDLKSIYKNSASKS 539
Db 480 SSRMELSCSLEQTOPAGPSLWKSKE-HQSONATLDREADTSSVGGMDIELNRKSPDRKP 538
QY 540 HAAEKLRSNKKREMDVVAIEDEVLEQLFKDTKPELEIDVVKQOEEDVNVKRPMDIET 599
Db 539 LPTETLPRKRKDV-DLATEEVEVLEELLRSTKPELAVQVKEQEAADTIRKKPRMDAER 597
QY 600 NDTFSDAEVPESSKISQENIEGKRELKEDSLWSAK-EISNNDKLODDSEMLPKLLLTE 658
Db 598 NRPLNGGSEPESSALQEDEREKKDELQTES-WSTRKHEIANSGLQDSEELPRKLLTE 656
QY 659 FRSLVTK--NSTSRNPSGINDGYQLKNFKKFKKVTYPGAGKLPHIIGSDLIHAHARKN 716
Db 657 FRSLVSNHNSTSRNLIC-VN-ECGPLKNFKFKKATFPGAGKLPHIIGSDLVGHARKN 714
QY 717 TELEEWLRQEMEVQONHAKESLADDLFRYNPYLKR 753
Db 715 TELEEWLKQEMEVQKQAKESLADDLFRYNPNVKKR 751

RESULT 4
Q9R1X1 PRELIMINARY; PRT; 751 AA.
ID Q9R1X1;
AC Q9R1X1;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE NIBRIN.
GN NBN OR NBS1.

```
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE-MIXTURE OF BRAIN, AND TESTIS;
RA Saito T.;
RT "Structure of the mouse Nijmegen breakage syndrome (Nbrin/Nbs1)
RL protein.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016988; BAA76298.1; -.
DR MGD; MGI:1351625; Nbn.
DR InterPro; IPR001357; BRC1.
DR InterPro; IPR00253; FHA_domain.
DR Pfam; PF00533; BRC1; 1.
DR Pfam; PF00498; FHA; 1.
DR SMART; SM00292; BRC1; 1.
DR SMART; SM00240; FHA; 1.
DR PROSITE; PS50006; FHA_DOMAIN; 1.
SQ SEQUENCE 751 AA; 83842 MW; ABA75A61F57838F9 CRC64;

Query Match 69.3%; Score 2702; DB 11; Length 751;
Best Local Similarity 70.7%; Pred. No. 1.7e-150;
Matches 535; Conservative 82; Mismatches 130; Indels 10; Gaps 9;

QY 1 MWKLLPAAGPAGGPPYRLLTGVEYVVGKNCALIIENDQSISRHHAVLTANFVNLSQT 60
DB 1 MWKLLPAAGPAGGPPYRLLTGVEYVVGKNCALIIENDQSISRHHAVLTANFVNLSQT 60
QY 61 DEIPVLTKDMSKYGTFFNEEKMONGSRITKSGDITFGVFGSKFRYEYPLVACSCL 120
DB 61 DEIPVLTKDMSKYGTFFNEEKMONGSRITKSGDITFGVFGSKFRYEYPLVACSCL 120
QY 121 DVSGKTALNQAILQGGFTVNNWTEECTHLVWVSVKVTIKTICALICGRPIVKPEYFTEF 180
DB 121 DVSGKTALNQAILQGGFTVNNWTEECTHLVWVSVKVTIKTICALICGRPIVKPEYFTEF 180
QY 181 LKAVOSKKQPPQIESFYPPIDEPALGSKSVDLGRHQRKQIFKCKTFVFLNAKHKLSS 240
DB 181 LKAVOSKKQPPQIESFYPPIDEPALGSKSVDLGRHQRKQIFKCKTFVFLNAKHKLSS 240
QY 241 AVFEGGGEARLMAEDDEEEOQSFAPGTCVVDGVTITQLIISHSQKWIHLIMDTLQRN 300
DB 241 AVFEGGGEARLMAEDDEEEOQSFAPGTCVVDGVTITQLIISHSQKWIHLIMDTLQRN 300
QY 301 GLRPIPAEIGLAVIFMTTKNYCDPQGHSTGLTKTTTPGSLSGVSGVDEKLMPSAPVNT 360
DB 301 GLRPIPAEIGLAVIFMTTKNYCDPQGHSTGLTKTTTPGSLSGVSGVDEKLMPSAPVNT 360
QY 361 TTYVADTSEQADT-WDLSEPRKEIKVSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVSN 419
DB 361 TTYVADTSEQADT-WDLSEPRKEIKVSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVSN 419
QY 420 TLAKMRIPNQLSTKPLPSINKSDRASQOQNTSIINYPQSTPKRERDEENQEMSSCK 479
DB 420 TLAKMRIPNQLSTKPLPSINKSDRASQOQNTSIINYPQSTPKRERDEENQEMSSCK 479
QY 480 SARLETSCLEQTPATPSLWKNKEOHLSENEPVDNTSDNLTFTDLSIKVNSAKS 539
DB 480 SARLETSCLEQTPATPSLWKNKEOHLSENEPVDNTSDNLTFTDLSIKVNSAKS 539
QY 480 SSRMELSCLEQTPAGPSLWSKE-HSQSNATLDREADTSSVGGMDIELNRKSPDRKP 538
DB 480 SSRMELSCLEQTPAGPSLWSKE-HSQSNATLDREADTSSVGGMDIELNRKSPDRKP 538
QY 540 HAAEKLRSKREMDVVAIEVLEQLFKDTKPELEIDVQKOEEDVNVKRPRMDIET 599
DB 540 HAAEKLRSKREMDVVAIEVLEQLFKDTKPELEIDVQKOEEDVNVKRPRMDIET 599
QY 539 LPTETLPRKRKDV-DLATEEVELEELRSTKPELAVQVKEQADDTIRKPRMDAER 597
DB 539 LPTETLPRKRKDV-DLATEEVELEELRSTKPELAVQVKEQADDTIRKPRMDAER 597
QY 600 NDTFSDAEVPESSKISENEIGKRELKEDSLASAK-EISNNDKLQDDSEMLPKLLLTE 658
DB 600 NDTFSDAEVPESSKISENEIGKRELKEDSLASAK-EISNNDKLQDDSEMLPKLLLTE 658
QY 598 NRPLNGSEPSALQEDEREKDELQTES-WSTKHEIANSGLQDSSELPKRLLTE 656
DB 598 NRPLNGSEPSALQEDEREKDELQTES-WSTKHEIANSGLQDSSELPKRLLTE 656
QY 659 FRSLIVK--NSTSRNPISGINDYGLQKNFKFKVTPYAGAKLPHIIGGSDLLTAHARKN 716
DB 659 FRSLIVK--NSTSRNPISGINDYGLQKNFKFKVTPYAGAKLPHIIGGSDLLTAHARKN 716
```

```
Db 657 FRSLVSVNHNSTSRNL-VN-EGGLKKNFKKATFFCAGKPLHIIGGSDLVGHARKN 714
QY 717 TELEEWLRQEMEVQNHAKESLADDLFRYNPNYKRR 753
DB 715 TELEEWLRQEMEVQNHAKESLADDLFRYNPNYKRR 751

RESULT 5
QY 09JIL9 PRELIMINARY; PRT; 750 AA.
AC 09JIL9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NBS1.
GN NBS1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20368653; PubMed=10908350;
RA Lanson N.A. Jr.; Egeland D.B.; Royals B.A.; Claycomb W.C.;
RT "The MRE11-NBS1-RAD50 pathway is perturbed in SV40 large T antigen-
RT immortalized At-1, At-2 and HL-1 cardiomyocytes.";
RL Nucleic Acids Res. 28:2882-2892(2000).
DR EMBL; AF218575; AAF91228.1; -.
DR InterPro; IPR001357; BRC1.
DR InterPro; IPR000253; FHA_domain.
DR Pfam; PF00533; BRC1; 1.
DR Pfam; PF00498; FHA; 1.
DR SMART; SM00292; BRC1; 1.
DR SMART; SM00240; FHA; 1.
DR PROSITE; PS50006; FHA_DOMAIN; 1.
SQ SEQUENCE 750 AA; 83304 MW; 7368BAD0914CF305 CRC64;
```

```
Query Match 68.9%; Score 2685.5; DB 11; Length 750;
Best Local Similarity 69.5%; Pred. No. 1.6e-149;
Matches 525; Conservative 93; Mismatches 130; Indels 7; Gaps 7;

QY 1 MWKLLPAAGPAGGPPYRLLTGVEYVVGKNCALIIENDQSISRHHAVLTANFVNLSQT 60
DB 1 MWKLLPAAGPAGGPPYRLLTGVEYVVGKNCALIIENDQSISRHHAVLTANFVNLSQT 60
QY 61 DEIPVLTKDMSKYGTFFNEEKMONGSRITKSGDITFGVFGSKFRYEYPLVACSCL 120
DB 61 DEIPVLTKDMSKYGTFFNEEKMONGSRITKSGDITFGVFGSKFRYEYPLVACSCL 120
QY 121 DVSGKTALNQAILQGGFTVNNWTEECTHLVWVSVKVTIKTICALICGRPIVKPEYFTEF 180
DB 121 DVSGKTALNQAILQGGFTVNNWTEECTHLVWVSVKVTIKTICALICGRPIVKPEYFTEF 180
QY 181 LKAVOSKKQPPQIESFYPPIDEPALGSKSVDLGRHQRKQIFKCKTFVFLNAKHKLSS 240
DB 181 LKAVOSKKQPPQIESFYPPIDEPALGSKSVDLGRHQRKQIFKCKTFVFLNAKHKLSS 240
QY 241 AVFEGGGEARLMAEDDEEEOQSFAPGTCVVDGVTITQLIISHSQKWIHLIMDTLQRN 300
DB 241 AVFEGGGEARLMAEDDEEEOQSFAPGTCVVDGVTITQLIISHSQKWIHLIMDTLQRN 300
QY 301 GLRPIPAEIGLAVIFMTTKNYCDPQGHSTGLTKTTTPGSLSGVSGVDEKLMPSAPVNT 360
DB 301 GLRPIPAEIGLAVIFMTTKNYCDPQGHSTGLTKTTTPGSLSGVSGVDEKLMPSAPVNT 360
QY 361 TTYVADTSEQADT-WDLSEPRKEIKVSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVSN 419
DB 361 TTYVADTSEQADT-WDLSEPRKEIKVSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVSN 419
QY 420 TLAKMRIPNQLSTKPLPSINKSDRASQOQNTSIINYPQSTPKRERDEENQEMSSCK 479
DB 420 TLAKMRIPNQLSTKPLPSINKSDRASQOQNTSIINYPQSTPKRERDEENQEMSSCK 479
QY 421 TLVRGKAPNQLSPMKCPAAKKNKDWSS-OQQLNSIRKNYFQPSKRERDEENQESCK 479
DB 421 TLVRGKAPNQLSPMKCPAAKKNKDWSS-OQQLNSIRKNYFQPSKRERDEENQESCK 479
```

| | | | |
|--------|---|--|--------------|
| Qy | 480 | SARIETSCSLLLEQTQATPSLWKNKQHLSENEPVDNTSDNNLFTDTDLKLSIVKNSAKS | 539 |
| Db | 480 | SSRVLLSCSLLLEQTQATPSLWKNKQHLSENEPVDNTSDNNLFTDTDLKLSIVKNSAKS | 538 |
| Qy | 540 | HAAEKLRSNKKRREMDVAIDEVLEOLFQDTPLELIDVYKQOEADVNVKRPRMDIET | 599 |
| Db | 539 | FSTEDLRARARKEV-DLSTEEVELELLRSTPELAVQVYKVKQOEADVNVKRPRMDIET | 597 |
| Qy | 600 | NDTFSEADVAPESSKISQENEIGKRRKELKEDSLWSAK-EISNNDKLQDDSEMLPKKLLLTE | 658 |
| Db | 598 | NQHLGGPVPESNSALQEDETEKKDELQTEA-WSTKREVSNTDELQDSSEELPRKLLTE | 656 |
| Qy | 659 | FRSLVTKNSTSRNPISGINDYQGLKNFKFKVTPYCGAKLPHIITGGSDLIHAHARKNTE | 718 |
| Db | 657 | FRSLVHNHNSRNLCLPLN-GRGELKNFKFKKATCPGAGKLPHIITGGSDLIHGHAARKNTE | 715 |
| Qy | 719 | LEWLROEMEVQNOHAKESLADDLFRYNPLKRR | 753 |
| Db | 716 | LEWLKHMEVQKQAKEDSLADDLFRYNPNVKRR | 750 |
| RESULT | 6 | | |
| ID | Q9DE07 | PRELIMINARY; | PRT; 753 AA. |
| AC | Q9DE07; | | |
| DT | 01-MAR-2001 | (TEMBLrel. 16, Created) | |
| DT | 01-MAR-2001 | (TEMBLrel. 16, Last sequence update) | |
| DT | 01-DEC-2001 | (TEMBLrel. 19, Last annotation update) | |
| DE | NIJMEGEN BREAKAGE SYNDROME 1. | | |
| GN | NBS1. | | |
| OS | Gallus gallus (Chicken). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; | | |
| OC | Gallus. | | |
| OX | NCBI_TaxID=9031; | | |
| ON | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE:20576328; PubMed:11062235; | | |
| RA | Tauchi H., Kobayashi J., Morishima, Ki, Matsuura S., Nakamura A., | | |
| RA | Shiraishi T., Ito E., Masnada D., Delia D., Komatsu K.; | | |
| RT | "The Forkhead-associated Domain of NBS1 Is Essential for Nuclear Foci | | |
| RT | Formation after Irradiation but Not Essential for hRAD50.hMRE11.NBS1 | | |
| RT | Complex DNA Repair Activity."; | | |
| RL | J. Biol. Chem. 276:12-15(2001). | | |
| DR | ENBL; AF230342; AAC47947.1; -. | | |
| DR | InterPro: IPR001357; BRCT. | | |
| DR | InterPro: IPR000253; FHA_domain. | | |
| DR | Pfam: PF00533; BRCT; 2. | | |
| DR | Pfam: PF00498; FHA; 1. | | |
| DR | SMART: SM00292; BRCT; 1. | | |
| DR | SMART: SM00240; FHA; 1. | | |
| DR | PROSITE: PS50006; FHA_DOMAIN; 1. | | |
| FT | VARIANT 250 250 | A -> T. | |
| FT | VARIANT 285 285 | W -> R. | |
| SQ | SEQUENCE 753 AA; 83986 MW; 410BBE74123D9B06 | CRC64; | |

| | | | | |
|-----------------------|-------|--|----------------|------------|
| Query Match | 45.3% | Score 1766.5 | DB 13 | Length 753 |
| Best Local Similarity | 49.4% | Pred. No. 1.1e-95 | | |
| Matches | 385 | Conservative 135 | Mismatches 207 | Indels 53 |
| Gaps | 20 | | | |
| Qy | 1 | MWKLPAAGPAGGEPYRLLTCTVEYVVGKNCAILIENDOSISRNHAVLTANFSTVNL | SQT 60 | |
| | | | | |
| Db | 1 | MWKLVPAAGP--GEPFLLVGTVEYVVGKNCACFLIQDDQISRHAVLTVRPETH | SQS 58 | |
| | | | | |
| Qy | 61 | DEIPVLTLKNSKYGTTFVNEEKMGFSSRTLKSDGDTTFGVGSKFEYEPLVAC | SCL 120 | |
| | | | | |
| Db | 59 | VSPVPLIKTSKYGTTFVNGSKL-SGASRSIQSDGRVNFVGFESKFEYESLV | CSSCL 117 | |
| | | | | |
| Qy | 121 | DVSKKTALNQAILOLGFTVNNWTTECHTLVMVSKVTKITKICALICGRPIVK | PEYTFEF 180 | |
| | | | | |
| Db | 118 | DVAOKTALNEAIQOLGLVNWETKECTHLLMESVTKVTKICALICGRPIVK | PEFSSEL 177 | |
| | | | | |

| | | | |
|--------|---|--|---------|
| QY | 181 | LKAVOSKKQPQIESFYPPIDEPSIGSKNVLDLSGRQERKQIFKGKTFIFILNAKQHKHKLSS | 240 |
| Db | 178 | MKAVOSRQQLPTPTSEFYPSPDEPAIGDNDMLSGHPHKKIFSGKTFVLTKAQHKHKLGP | 237 |
| QY | 241 | AVVFGGGEARLIITENEENEHNFPLAGCTGVVDVGIINSQTLIPDCOKKWIQSIQIMDLQRO | 300 |
| Db | 238 | AVILVGGGEAKLMAERKET-SLIVSPVCVVDGVGTVNSQLGSESRNNTWDSILAVLESN | 296 |
| QY | 301 | GLRPIPEAIEGLAVIFMTTKNYCDPOGHPSTGLKTTTPG----PSLSQGVSDVEKILMPSA | 356 |
| Db | 297 | NLRRAIPEAIEGLAVIFMSTEYICNPQRQPDNKAVTASTASKVRPVSQSSTVDETIMPTA | 356 |
| QY | 357 | PVN-TTTTVADTSEQADTWDLSEKPEIKVSKMEOKFRLMSODATPTVKECKTSSNNNS | 415 |
| Db | 357 | AADYSTLVNADTTEIEQTCMEI-----ERTTSQTTREKVAFOQA-AVRENPFSTSGTVNA | 410 |
| QY | 416 | -MYS--NTLAKMRIPNYQLSPTKLPSINKSDRASQOQOTNSIRNYFOPSTKKRERDEE | 471 |
| Db | 411 | GMLISRVNRTSGFGQKNHPSKILIEVDKPRE-CPTRQOSNSITNFHVA-RKRERAE | 468 |
| QY | 472 | NOEMSSCKSARIETS-CSLLEQOTQATPPLSWKN-KEOHLSENEPVDNDSNNLFTD-TDL | 528 |
| Db | 469 | GEETSLSQAKLEKKPLPVSECTESSASSAWNSEKEQHGKGNNIQLGRESGELASDKTDI | 528 |
| QY | 529 | K-SIVKNSAKSHAAEKLRNKKREMDVAIEDVEULQFKDTKPELEIDVKVQKQEDV | 587 |
| Db | 529 | KITFSENPAKP-----KKELDVSDEYETLEWFESE-R-DLDWEEQTANGDQEA | 576 |
| QY | 588 | NVRKPRMDTETNDTFDEAVPESSKISQSENEIGKKRE-----LKEDSLWSAK | 635 |
| Db | 577 | QSNKRRKRLKETGSRTEEG---NTKORENEMLRKEEVGSVLTLDEKSKIRESSVSIR | 633 |
| QY | 636 | -EISNNDKLODDSEMLPKLLLTFFRSLVTKNSTSRNPSPGINDDYQGLKNFKFKKVTYP | 694 |
| Db | 634 | NKLINHKNLEDDSRPLSKLLLTFFRSLVVSVCPSRNSPTMRNTKCRGNQNFKTRKVPVP | 693 |
| QY | 695 | GAGKLPHIIGSDLIHAHAKNTELEBWLKQEMVQNOHAKESLADDLFRNYPYLKRRR | 754 |
| Db | 694 | GAGOLPYIIGSDLVHAHQARKNSELEBWLKEELEQNRRAEESLADDLFRYDPNVKRRR | 753 |
| RESULT | 7 | | |
| Q9VT40 | | | |
| ID | Q9VT40 | PRELIMINARY; | 811 AA. |
| AC | Q9VT40; | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Created) | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Last sequence update) | | |
| DT | 01-JUN-2001 (TrEMBLrel. 17, Last annotation update) | | |
| DE | PUTATIVE NBS PROTEIN. | | |
| OS | NBS OR CG6734. | | |
| GN | Drosophila melanogaster (Fruit fly). | | |
| OC | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | | |
| OC | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | |
| OC | Ephydroidea; Drosophilidae; Drosophila. | | |
| NCBI | NCBI_TaxId=7227; | | |
| RN | [1] | | |
| RC | SEQUENCE FROM N.A. | | |
| RP | STRAIN=BERKELEY; | | |
| RX | MEDLINE=20196006; PubMed=10731132; | | |
| RX | Adams M.D., Celisnik S.E., Holt R.A., Evans C.A., Gocayne J.D., | | |
| RA | Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., | | |
| RA | George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., | | |
| RA | Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., | | |
| RA | Randall R.C., Rogers Y.-H.C., Blazek R.G., Champé M., Pfeiffer B.D., | | |
| RA | Wan K.H., Doyle C., Baxter E.G., Held G., Nelson C.R., Miklos G.L.G., | | |
| RA | Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., | | |
| RA | Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., | | |
| RA | Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., | | |
| RA | Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P., | | |
| RA | Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., | | |
| RA | Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., | | |
| RA | de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., | | |

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Metkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weltsenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of Drosophila melanogaster.";
 RT Science 287:2185-2195(2000).
 RL EMBL; AE003551; AAF50215.1; -.
 DR FlyBase; FBgn0026198; nbs.
 DR InterPro; IPR001357; BRCT.
 DR InterPro; IPR00253; FHA_domain.
 DR Pfam; PF00498; FHA; 1.
 DR SMART; SM00292; BRCT; 1.
 DR SMART; SM00240; FHA; 1.
 DR PROSITE; PS50006; FHA_DOMAIN; 1.
 DR KW Hypothetical protein.
 SQ SEQUENCE 811 AA; 90286 MW; 2BB9936468B05E8E CRC64;

Query Match 12.1%; Score 473; DB 5; Length 811;
 Best Local Similarity 23.8%; Pred. No. 8.3e-20;
 Matches 199; Conservative 129; Mismatches 297; Indels 210; Gaps 31;

Qy 8 AGPAGGEPYRLLTGVE-VVVGKNCALLIENDQSTSRNHAVITANFSVNLISQTDIPLV 66
 Db 17 AAIADEKFLVLPKPKVITIGRLATDLIVAQDLISRNHAQLIQ-----TEADGGDTL 70
 Qy 67 TLKD-NSKYGTGV---NEEKMGFSRT---LKSGDGTTFGVGSKFRIEYPLVACSSC 119
 Db 71 HIEDLSRYGTFIPKNSQKPRKVPKSTPLPVGTRLRFGANMSIWQVTLKLVTVSA 130
 Qy 120 LDVSGKTALNAIQLGGFTVNNWTECTHLVMVSVKVTIKTICALICGRPIVPEYFTE 179
 Db 131 LTRSEVQELTKMLPEPMGGTGTSTNWTCEGSHLTMTNEVSVTVKLLHMLNKKPIVTFPYWRK 190
 Qy 180 FLKAVQS---KKPOPIESYPPLDPEISGSKNVDLSGRQERKQIFKGTIFLNAKQHK 236
 Db 191 MLQAAQSLHVKEGWPQEDYQ-----TNIDVTWRPRTLFAKGTIVEMNRKHF 241
 Qy 237 KLSAVVFGGGEARLITENEENHFFLAPGTCVVDITGITSQTLIPDCQKKWQISMDM 296
 Db 242 MYGVSQKAGATCRDI---NSGVRKFTLTKSDVIV-----IQVPSQSQATESINNI 291
 Qy 297 LQROGLRPIPAEIGLAVIFMTKNYCDPOGHP--STGLKTTTPGPSISQGVSVDEKLM 354
 Db 292 LEQNGRRIIQIEYIGMALIHCSITEFCNPT-HKFISDSLPT-----ESVTS 337
 Qy 355 SAPVNTTYYVADTE--SQADTWLSEK--PKEIKVSKWEQKFMLSODAPTVKE--SK 408
 Db 338 SMARNSSIIVPTEHSAQSNATPISLVVPESECEMEQDASPHSDQASRRSHAS 397
 Qy 409 TSSNNNSMVSNTLAK-----MRIPN-----YQLSPTKLPS-----INKSK 443
 Db 398 TVDSSDEBKSTLAKRAKSDIATKLTMTKSKNAILLDSLEEDVTAPAPAPQVTRQSK 457

Qy 444 DRASQQOQNTSIRNFQPSQTKKRER-----DEENQEMSKSARIETSCSLLEQTOPA 496
 Db 458 AIAEKSXVHPVPAASKHITRKTQVFCVDSDEENENARKPK-----ETPAPT 506
 Qy 497 TPSLWKNEQHLSENEPVDNTSDNNLFTDTDLKSLTVKNSASKSAAEKLRSNKKREMDIV 556
 Db 507 IPSMAKKKTE-----APVATRISPRL-NGKSLATNITNPADKHAVP-----AKRPVLVS 555
 Qy 557 AIEDVLE-QLFKDTK-PELEIDVKVQ-----SGINDD---YGOLKNFKFKKVTYP 581
 Db 556 ASSDEEDGDLFQPKSPQKPAETVVQPRIAGKNAPARISVVDFLEKSQAQEPAPVPQ 615
 Qy 582 -KQEDVNVNRKPRMDIETNDFTSDEAVP-----ESSKISQENEIGKKRELKEDSLMSAKE 636
 Db 616 LESQSQTPRKRLRLEL-LNESDSDCCNLFNFADSKKKRKTQEAQRNDDSDTGLFNFS 674
 Qy 637 ISNNDKLQDDSEML-----PKLLLLTEFSLIVIKNSTRNP- 672
 Db 675 ERPSDHDDEDSRLTEPFPVPETESKKQSKYIVAPRRDRPKKVDISDADSVMKETSIAKADP 734
 Qy 673 -----SGINDD---YGOLKNFKFKKVTYP 694
 Db 735 EEQWLAAMKDSIEVRMCNLNIVIRSQEVDASLEDSVKNHGGKKNFKFVKTKNP 799
 RESULT 8
 ID 075058 PRELIMINARY; PRT; 1460 AA.
 AC 075058;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE KIAA0470 PROTEIN.
 GN KIAA0470 OR KAB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RC MEDLINE=98116662; PubMed=9455484;
 RA Seki N., Ohira M., Nagase T., Ishikawa K., Miyajima N., Nakajima D.,
 RA Nomura N., Ohara O.;
 RT "Characterization of cDNA clones in size-fractionated cDNA libraries
 RT from human brain.";
 RL DNA Res. 4:345-349(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Hara Y., Adachi Y.;
 RT "Molecular cloning and initial characterization of KAB.";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB007939; BAA32315.1; -.
 DR EMBL; AB022659; BAA83380.1; -.
 DR InterPro; IPR000253; FHA_domain.
 DR Pfam; PF00498; FHA; 1.
 DR SMART; SM00240; FHA; 1.
 DR PROSITE; PS50006; FHA_DOMAIN; 1.
 SQ SEQUENCE 1460 AA; 161436 MW; BAD23EBCA19B65F0 CRC64;

Query Match 5.8%; Score 227.5; DB 4; Length 1460;
 Best Local Similarity 18.8%; Pred. No. 4.2e-05;
 Matches 154; Conservative 149; Mismatches 318; Indels 199; Gaps 29;

Qy 2 WKLLPAAGPAGEPYRLLTGVEYVVGVGKNCALLIENDQSTSRNHAVITANFSVNLSTQD 61
 Db 6 WFLVSSGGTRHLRPREMI-----FVGRDDCELMQLQ-RSDVKQHAVINYNAS-----TD 53
 Qy 62 EIPVLTKDNKSYGTVEENKMQGFSRTLSKSGDITGVGSKFRIEYPLVACSSCLD 121
 Db 54 EHLVKDL--GSINGTFVNDVRPIEQTYITLKLKDLRFQYDNTLFTV----- 98


```
QY 662 LVIKNSTSRNPSGI-----NDYGO-----LKNFKFKKVVTPGAGKLPHIIGGSDLIAH 711
: ||| | | : | : ||| |
Db 1303 ATWNLNIIIFASGIATREENTDVLVEERIQRVEFKTT-----1343

QY 712 HARKNTELE-EWLQRQEMEQVONQHAKEESLA-----DDLFF 744
: | | : | : | | | | : | : | |
Db 1344 ---ENLEIQKEVVLTKEEVDNSDKVKEHRTSAVNIDLDVDF 1380

RESULT 15
P91257
ID P91257 PRELIMINARY; PRT; 3484 AA.
AC P91257;
DT 01-MAY-1997 (TremBLrel. 03, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE HYPOTHETICAL 385.7 KDA PROTEIN.
GN F12F3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Fulton B., Wohlmann P.;
RT "The sequence of C. elegans cosmid F12F3.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U80022; AAC25885.2; -.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR003962; FnIII_repeat.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_Like.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00041; fn3; 7.
DR Pfam; PF00047; ig; 10.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM00060; FN3; 8.
DR SMART; SM00409; IG; 12.
DR SMART; SM00408; Igc2; 8.
DR SMART; SM00410; IG_Like; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Hypothetical protein; Immunoglobulin domain; Repeat.
SQ SEQUENCE 3484 AA; 385746 MW; 21DD6F6893020F4F CRC64;
```

```
Query Match 4.8%; Score 185.5; DB 5; Length 3484;
Best Local Similarity 20.0%; Pred. No. 0.035;
Matches 132; Conservative 114; Mismatches 237; Indels 177; Gaps 31;

QY 166 ICGRPVKEYTEFLKAVQSKK---OPQIESFYPLDPEPSIGSKNVDLSGQERKQIF 222
: | | : | | : | | : | | : | |
Db 466 VTGKPL-----EAKKPVEDKKDASQSSSKESPPTD-----KKQIP 505
```

```
QY 223 KGKTFIFL-----NAQHKHKLSSAVVFGGE-----ARLITEEN 256
: | | : | : | | : | |
Db 506 KA---LFIPDEISSRFCDPSTMHSETNITTTIRREGSADAKTLPVPLSASVMKVESA 562

QY 257 EEEHNFELAPGTCVVDGTINSOTLIPDCQKWKIQSMDMLQROGLRPIPEAEIGLAVIF 316
: | : | | | : | : | : | : | : | : |
Db 563 KEAEFSFK-----RRSET--PDDKSR-----KKEGLPPAKKSE-----594

QY 317 MITKNYCDPQGHPSGTGLKTTTPGPSLSQGVSDVE-KILMPSAP-----VNTTIVYADT 367
: | | : | | : | | : | | : | |
Db 595 -----KKDEVTAEKQSTEALESKKKEVDEKISEQOPSDKNKSEVGVPERAAGP 645

QY 368 ESEQADTWLDSERPEIKVSKMEQKF-RMLSQDAPTVKESCKTSSNNNSVSTLAKMRI 426
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 646 ETKK-DVSEIEEVPKKTIKKTEKSDSSISQKSNVLKPADDDKSKSDDDVTDKSKKTT- 703

QY 427 PNQLSPTKLPSINKSKDRASQOQOTNSIRNYFPQSTKKRERDEENQEMSSCKSARLET 486
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 704 -----DQTKVATDSKLEKAADTKQIET-ETVVDDKSKKVKLKKKTEKSDSFISQKSETP 757

QY 487 CSLEQOTQATPSLWKNKEQHLSENEPVDNNDNNLFTDLDKSIYKNKSASKSHAAEKL 546
: | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 758 -PVVEPTKPAESEAQKIAEVNKAQKQ---KEVDNLKREAEV-----AAKKIADEKLK 806

QY 547 -----SNKKREDD-VAIEDEVLEQLFKDTKPELEIDVKVQKOEEDVNR 590
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 807 IEAEANIKKTAEEVAARKQEKDEQLKLETEVYVSKSAAEKLELEKQAQIKKAAEADAVK 866

QY 591 KRPRMDIETND-----TFSDAEVPESSKISOEN-EIGKKRELK-----ED 629
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 867 KQKELN-EKNKLEAAKSAADKLLKEESAASKKSESVKFGEEKKTAGEKTVQVES 925

QY 630 SLWSAKEISNND-KLQDDSEMLPKLLLTETFRSLVKNSTSRNPNGIND----DYGLKN 684
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 926 EPTSCKTTIDTKDVGATEPADETPKK-----KIKKKTEKSDSSISQKSDATSEKSVK 977

QY 685 FKFKKVTYTPGAGKLPHIIGGSDLIAHHARKNTELEWLRQEMEV---QNHAKESLAD 741
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 978 QKEQDEPTKPAVSETQMTVE-----ADKSKKQKETDEKLDKDAEIAAKTKQEADEKSKLD 1032
```

Search completed: August 15, 2002, 01:48:48
Job time: 488 sec

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|-----------|---------------------|
| 1 | 4105.2 | 93.2 | 4386 | 21 | AAZ890047 | Human nibrin DNA. |
| 2 | 4103.6 | 93.2 | 4386 | 21 | AAZ324997 | NBS1 gene associat |
| 3 | 1862.8 | 42.3 | 65921 | 22 | AAZ890046 | Human nibrin DNA. |
| 4 | 1836.4 | 41.7 | 2044 | 22 | AAH13813 | Human CDNA sequenc |
| 5 | 871 | 15.2 | 752 | 22 | AAH03660 | Human CDNA clone (|
| C 6 | 552.2 | 12.5 | 646 | 21 | AAAT78036 | CDNA encoding huma |
| C 7 | 552.2 | 12.5 | 646 | 22 | AAI28794 | Colon tumour relat |
| C 8 | 520.4 | 11.8 | 543 | 22 | AAH09188 | Human CDNA clone (|
| 9 | 358.4 | 8.1 | 396 | 22 | AAF44923 | Human breast cancer |

SUMMARIES

DR P-PSDB; AAY51669.

XX A DNA double strand break repair protein, Nibrin, and related DNA
PT useful for diagnosis and therapy of Nijmegen Breakage Syndrome and
PT other diseases influenced by DNA-double-strand break repair -
XX
XX Claim 2a; Fig 2; 32pp; German.
XX
CC This invention describes a novel DNA double strand break repair protein,
CC Nibrin. Nibrin and DNA encoding it are useful for diagnosis and/or
CC therapy of diseases influenced by repair of DNA-double strand breaks,
CC in particular Nijmegen Breakage Syndrome. The product of the invention
CC has applications in gene therapy. This sequence encodes the nibrin
CC protein described in the invention.
XX
SQ Sequence 4386 BP; 1467 A; 744 C; 845 G; 1330 T; 0 other;

Query Match 93.2%; Score 4105.2; DB 21; Length 4386;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 4270; Conservative 7; Mismatches 30; Indels 91; Gaps 6;
QY 28 gccccagccctgagagccgacccgatgtgaaactctgccccgcccggccgagc 87
DB 1 gccccagccctgagagccgacccgatgtgaaactctgccccgcccggccgagc 60
QY 88 agggagaaccatacacagacttttgactggcgttgtagtgcgtttgtgaaaggaaactgtgc 147
DB 61 agggagaaccatacacagacttttgactggcgttgtagtgcgtttgtgaaaggaaactgtgc 120
QY 148 catctaatgaaatgatacagtcagtcagccgaaatcatgctgtttaactgttaactt 207
DB 121 catctaatgaaatgatacagtcagtcagccgaaatcatgctgtttaactgttaactt 180
QY 208 ttctgttaaccaactgagtcacacagatgaatccctgtattgacattaaagataatc 267
DB 181 ttctgttaaccaactgagtcacacagatgaatccctgtattgacattaaagataatc 240
QY 268 taagtgtgtacacctttgtaatgaggaaaaaatcagaatggcgtttcccgaaactttgaa 327
DB 241 taagtgtgtacacctttgtaatgaggaaaaaatcagaatggcgtttcccgaaactttgaa 300
QY 328 gtcggggagatgtattacttttgagtggtttggaagttaattcagaatagatagacc 387
DB 301 gtcggggagatgtattacttttgagtggtttggaagttaattcagaatagatagacc 360
QY 388 ttgtgttcagtcctctctgttttagatgtctctggaaaaactgctttaaatcaagctat 447
DB 361 ttgtgttcagtcctctctgttttagatgtctctggaaaaactgctttaaatcaagctat 420
QY 448 attgcaacttggagatttactgttaacaaattggacagaagaatgcactcaccttgcac 507
DB 421 attgcaacttggagatttactgttaacaaattggacagaagaatgcactcaccttgcac 480
QY 508 ggtatcagtgaaagtaccattaaacaaatattgtgcactcatttggagcgtcccaattgt 567
DB 481 ggtatcagtgaaagtaccattaaacaaatattgtgcactcatttggagcgtcccaattgt 540
QY 568 aaagccagaatatttactgaattccctgaaagcaggttcagtcacgaagcagcctccaca 627
DB 541 aaagccagaatatttactgaattccctgaaagcaggttcagtcacgaagcagcctccaca 600
QY 628 aattgaaagtgtttaccacacctttgatgaacctctatttggaaagtataaattgttgatct 687
DB 601 aattgaaagtgtttaccacacctttgatgaacctctatttggaaagtataaattgttgatct 660
QY 688 gtcaggagccgagaaagaaacaaatcttcaagggaacacatttatatttttgaatgc 747
DB 661 gtcaggagccgagaaagaaacaaatcttcaagggaacacatttatatttttgaatgc 720
QY 748 caaacagcataaagaaattgagttcccgagttgtctcttggaggtgggaagcagtagttgat 807
DB 721 caaacagcataaagaaattgagttcccgagttgtctcttggaggtgggaagcagtagttgat 780

QY 808 aacagaagagaatgaagaagaacataatttttttggctccgggaacgtgtgttttga 867
DB 781 aacagaagagaatgaagaagaacataatttttttggctccgggaacgtgtgttttga 840
QY 868 tacagggaataacaaactcacagacacttaattcctgactgtcagaagaatgattcagtc 927
DB 841 tacagggaataacaaactcacagacacttaattcctgactgtcagaagaatgattcagtc 900
QY 928 aataatgatatgtccaaaggcgaaggtcttagacctattctgaagcagaataatggatt 987
DB 901 aataatgatatgtccaaaggcgaaggtcttagacctattctgaagcagaataatggatt 960
QY 988 ggcggtgatttcatgactacaaagaattactgtgatcctcagggccatccagtcacagg 1047
DB 961 ggcggtgatttcatgactacaaagaattactgtgatcctcagggccatccagtcacagg 1020
QY 1048 attaaagacaaacactccaggaccgaagcctttccaaaggcgtgtcagttgtgataaaact 1107
DB 1021 attaaagacaaacactccaggaccgaagcctttccaaaggcgtgtcagttgtgataaaact 1080
QY 1108 aatgccagcgcccgagtgaaactacaaactacacatacgtagctgacacagaatcagagcaagc 1167
DB 1081 aatgccagcgcccgagtgaaactacaaactacacatacgtagctgacacagaatcagagcaagc 1140
QY 1168 agatacatgggatttggatgaaaggccaaagaatacaaaagtcctccaaatggaaacaaa 1227
DB 1141 agatacatgggatttggatgaaaggccaaagaatacaaaagtcctccaaatggaaacaaa 1200
QY 1228 attcagaatgctttcacaagacgcccactgtgaaaggagtccttgcaaaaaaacgctctaa 1287
DB 1201 attcagaatgctttcacaagacgcccactgtgaaaggagtccttgcaaaaaaacgctctaa 1260
QY 1288 taataatagtagtgcatacaatactttggctaaagtgaagatgagaatcccaactactcagcttc 1347
DB 1261 taataatagtagtgcatacaatactttggctaaagtgaagatgagaatcccaactactcagcttc 1320
QY 1348 accaactaaattgcccagtagtaataaaagttaaagtaggcttctcagcagcagcagac 1407
DB 1321 accaactaaattgcccagtagtaataaaagttaaagtaggcttctcagcagcagcagac 1380
QY 1408 caactccatcagaacactactttcagccgtctaccaaaaaagggaagggaagaagaaa 1467
DB 1381 caactccatcagaacactactttcagccgtctaccaaaaaagggaagggaagaagaaa 1440
QY 1468 tcaagaatgtcttcacgaatacagcaagaatgaaacgtctgttctctctttttagaaca 1527
DB 1441 tcaagaatgtcttcacgaatacagcaagaatgaaacgtctgttctctctttttagaaca 1500
QY 1528 aacacaacctgtcacacctcattgttgaaaaataaaggagcagcatctatctgagaatga 1587
DB 1501 aacacaacctgtcacacctcattgttgaaaaataaaggagcagcatctatctgagaatga 1560
QY 1588 gctgtggacacaaactcagacaataacttatttacagatacacagatttataaactattgt 1647
DB 1561 gctgtggacacaaactcagacaataacttatttacagatacacagatttataaactattgt 1620
QY 1648 gaaaaattctgcccagtaaatctcgtgcagaaagcctaagatcaaatataaaaaaggga 1707
DB 1621 gaaaaattctgcccagtaaatctcgtgcagaaagcctaagatcaaatataaaaaaggga 1680
QY 1708 aatggatgagtggccatagaagatgaagtatttgaacagttattcgaaggacacaaaacc 1767
DB 1681 aatggatgagtggccatagaagatgaagtatttgaacagttattcgaaggacacaaaacc 1740
QY 1768 agaattgaaattgattgaaagttcaaaacaggaagaagatgataatgttagaaaaag 1827
DB 1741 agaattgaaattgattgaaagttcaaaacaggaagaagatgataatgttagaaaaag 1800
QY 1828 gcaagatgattgataaacaacaaatgacactttcagtgatgaagcagtcaccagaaagttag 1887
DB 1801 gcaagatgattgataaacaacaaatgacactttcagtgatgaagcagtcaccagaaagttag 1860
QY 1888 caaaatatctcaagaagaaatgaaattgggaagaaacgtgaaactcaagggaagcactcactatg 1947

CC protein described in the invention.

| XX | Sequence | 65921 BP; | 19501 A; | 11699 C; | 12463 G; | 22149 T; | 109 other; |
|----|-----------------------|--------------|---------------|------------|---------------|----------|-------------|
| SQ | Query Match | 42.3% | Score 1862.8; | DB 21; | Length 65921; | | |
| | Best Local Similarity | 94.1%; | Pred. No. 0; | | | | |
| | Matches 2024; | Conservative | 9; | Mismatches | 26; | Indels | 91; Gaps 6; |

Query Match 42.3%; Score 1862.8; DB 21; Length 65921;
Best Local Similarity 94.1%; Pred. No. 0;
Matches 2024; Conservative 9; Mismatches 26; Indels 91; G

Qy 2278 tcttttagatacaatccttatttaaaaggagaagataactgaggattttaaaaagaag
|||||

Db 58191 tctttgcagatacaatccttatttcaaaaggagaagataactgaggatttttaaaagaag

Qy 2338 ccattggaataaaccttccttagtaagcatctactctcaggcccaacaagggttatatgaatatata

58251 ccattgaaaaacttccttagtaagcatctactctcagccaacaaggttatatgaatatata

[illegible]

Qy 2458 caaaaacttttgattctcttttgcataatgtaacaaattttttgtvctgtatttttcaggagcttttgcacatttgc

Db 58371 caaaacttttgattctcttttgatatgtatacaattgtttgttctgttttcaggcgttttgtcattg

Qy 2518 catcttttttcatttttaaatggtgttttggttttattaaatagtttaatatagtcacagttc

Db 58431 catcttttttcatttttaaatggttttgtttattaaatagtttaatatagtcacagttcc

Qy 2578 aaattctaaatrtaagtaagtaaggaactaaagtcacccttcaccattgtcctagctt

Db 58491 aaaattctaaatgtacgtaaggtaaa-gactaaagtcaccccttccaccattgtccttagctb

Qy 2638 act-----
|||

Db 58550 acttgggtccccctcagaaaaaattcatgatactactcattcttcttatgaatctttccaggga

Qy 2641 ----tatttttaataaatttcctacacaaaatgatagcataaa
|||||
59010

atatacgaattctctacacacttgccttttatctgcttcaaaccttcctgtcccttctg
3670
tttgagtccttatccaattccctatttttaaataaatttctcacacaaaatgatagcataaa

[illegible]

Qy 2739 aatataatgttttttaataatatttttttcttttttccattatgctgtagtctttaccctaaactctgg

Db 58729 aaataaatgtttttaatatattttttttttccattatgctgtagttctaacctaaaactctggg

Qy 2799 tgatccaaacaaaatggcttcagtgggcagatgtcacctacatgttattctagtactag

Db 58789 tgatccaaacaaaatggcttcagtggtcagatgtcacctacatgttatcttagtactagg

Qy 2859 aaactgaagaccatgtggagacttcataaacatgggttttagttttcaccagaatggaaa

Db 58849 aaactgaagaccatgtggagacttcatcaaacatgggttttagttttccaccagaatgggaaaaa

Qy 2919 gacctgtaccccttttttggtggtcttactgagctgggtgggtgctctgttttgagcttatt

Db 58909 gacctgtacccccccttttttgggtggtcttactgagctgggtgggtgctgtctgttttggagcttat

Qy 2979 tagagtccttagtctttcctactataaagtagaaaaatggtgagattgtttctttttctacc

Db 58969 tagagtccctagtttttcctactttataaaagttagaaaatggtgagattgtttttctttttcttaccac

[illegible][illegible]

Db 59089 ttcaagtctgtgtttcaaaaaataattcatgtatccagcaagaaggaggttccaa

1. The first group of people who are not in the labor force are those who are not in the labor force for any reason. This group includes people who are not in the labor force because they are not in the labor force for any reason. This group includes people who are not in the labor force because they are not in the labor force for any reason.

|||||
Db 60219 cttttttccccataagaatccagtcattgtaacactgtttatgaagaattatcc 60278
Qy 4299 ttctcatataattacctgccaattagtaaaaaatacaataaccatim 4348
|||||
Db 60279 ttctcatataattacctgccaattagtaaaaaatacaataaccataa 60328

RESULT 4
AAH13813
ID AAH13813 standard; cDNA; 2044 BP.
AC AAH13813;
XX
XX 26-JUN-2001 (first entry)
DE Human cDNA sequence SEQ ID NO:10767.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
OS
PN EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
PR
XX 27-AUG-1999; 99JP-0300253.
PR
XX 11-JAN-2000; 2000JP-0118776.
PR
XX 02-MAY-2000; 2000JP-0183767.
PR
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Oka T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 10767; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC of the present invention.
XX
XX Sequence 2044 BP; 723 A; 384 C; 442 G; 495 T; 0 other;

Query Match 41.7%; Score 1836.4; DB 22; Length 2044;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 1976; Conservative 0; Mismatches 1; Indels 119; Gaps 2;

Qy 13 cgcggttgcaogtcgcccagccctgaggagccgagccgatggaagaaactgctgcgcg 72
Db 67 cgcggttgcaogtcgcccagccctgaggagccgagccgatggaagaaactgctgcgcg 126

Qy 73 cgcggttgcaogtcgcccagccctgaggagccctgaggagccgagccgatggaagaaactgctgcgcg 132
Db 127 cgcggttgcaogtcgcccagccctgaggagccctgaggagccgagccgatggaagaaactgctgcgcg 186

Qy 133 aaggaataactgtgccattcttaattgaaatgacagtcgacgcgcgagaaatcatctgt 192
Db 187 aaggaataactgtgccattcttaattgaaatgacagtcgacgcgcgagaaatcatctgt 246

Qy 193 gtaactgtaactttctgttaaccacccctgagtcgaacagatgaatccctgtattgac 252
Db 247 gtaactgtaactttctgttaaccacccctgagtcgaacagatgaatccctgtattgac 306

Qy 253 attaaagataattctaagtatgtaccttggtaagagaaataatgcagaatgcctt 312
Db 307 attaaagataattctaagtatgtaccttggtaagagaaataatgcagaatgcctt 366

Qy 313 ttcccgaaactttgaagtcgggggagtggtattacttttggagttgttggagtaaatcag 372
Db 367 ttcccgaaactttgaagtcgggggagtggtattacttttggagttgttggagtaaatcag 426

Qy 373 aatagatgagccttgggttgatgctctcttcttggatgctctctgggaaactgc 432
Db 427 aatagatgagccttgggttgatgctctcttcttggatgctctctgggaaactgc 486

Qy 433 tttaaatcaagctatatgtcaacttgaggtttactgttaacaaatggacagagaatg 492
Db 487 tttaaatcaagctatatgtcaacttgaggtttactgttaacaaatggacagagaatg 546

Qy 493 cactcaccttgcattggtatcagtgaaagtaccattaaaacaaatgtgcactcattg 552.
Db 547 cactcaccttgcattggtatcagtgaaagtaccattaaaacaaatgtgcactcattg 606

Qy 553 tggagctccaatgtataagccagaaatatttactgaattcctgaaagcagttcagccaa 612
Db 607 tggagctccaatgtataagccagaaatatttactgaattcctgaaagcagttcagccaa 666

Qy 613 gaagcagcctcccaaaattgaaagttttaccaccttgcagtaaccatctattggaag 672
Db 667 gaagcagcctcccaaaattgaaagttttaccaccttgcagtaaccatctattggaag 726

Qy 673 taaaaatgttgatctgcaggacgcccgggaaagaaacaaatcttcaaaagggaacatt 732
Db 727 taaaaatgttgatctgcaggacgcccgggaaagaaacaaatcttcaaaagggaacatt 786

Qy 733 tatattttgaaatgcacaaagataagaattgagttccgcagttcttcttgaggtgg 792
Db 787 tatattttgaaatgcacaaagataagaattgagttccgcagttcttcttgaggtgg 846

Qy 793 ggaagctaggttgataacagagaagaatgaagaagacataatttcttttggctccggg 852
Db 847 ggaagctaggttgataacagagaagaatgaagaagacataatttcttttggctccggg 906

Qy 853 aacgtgtgttgatagacaggaaatacaaaactcagacacctaattcctgactgacgaa 912
Db 907 aacgtgtgttgatagacaggaaatacaaaactcagacacctaattcctgactgacgaa 966

Qy 913 gaaatgattcagtcgaataatgataatgcctccaaagcaaggtcttagacctattctga 972
Db 967 gaaatgattcagtcgaataatgataatgcctccaaagcaaggtcttagacctattctga 1026

Qy 973 agcagaataatggatggcgggtgatttctcagactcacaagaatactgtgactcctcagg 1032
Db 1027 agcagaataatggatggcgggtgatttctcagactcacaagaatactgtgactcctcagg 1086


```
OS Homo sapiens.
XX EP1074617-A2.
XX
XX
XX PD 07-FEB-2001.
XX
XX
XX PF 28-JUL-2000; 2000EP-0116126.
XX
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX
XX PA (HELI-) HELIX RES INST.
XX
XX OTa T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX PS Claim 3; SEQ ID 6023; 2537pp + CD ROM; English.
XX
XX CC The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX SQ Sequence 543 BP; 128 A; 111 C; 83 G; 219 T; 2 other;

Query Match 11.8%; Score 520.4; DB 22; Length 543;
Best Local Similarity 98.3%; Pred. No. 5e-98;
Matches 535; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 1564 ggagcagcatctatcagaatgagcctgtggacacaaactcagacaataacttatttac 1623
DB 543 GNAGCGCTTCTATCTGAGATGAGCCTG-GGACCCAAACTCAGCCCAATACTTATTTC 485
QY 1624 agatacagattaaactctattgtgaaaaattctgccagtaaatctcatgtgcagaaaa 1683
DB 484 AGATCCAGATTTAAATCTATTGTGAAAAATCTGCCAGTAAATCTCATGCTGCAGAAAA 425
QY 1684 gctaagatcaataaaaaaggggaaatggatgattgtgtgccatagaaatgaagtattgga 1743
DB 424 GCTAAGATCAATNAAAAAGGAAATGGATGATGTGCCATAGAAAGTAAATATTGGA 365
QY 1744 acagtattcaaggacacaaacacagattagaaattgatgtgaaattcaaaacagga 1803
DB 364 ACATTTATTCNAGGACACAAANNCAGAGTTGATGTGAAGTTCAAAACACGGA 305

1804 ggaagatgtaaatgttagaaaaagcccaagatggtatagataagaaaaaatgacactttcag 1863
DB 304 GGAAGATGTCATCTTAGAAAAAGGCCAAGGATGGATATAGAAACAATGACACTTTTCAG 245
QY 1864 tgatgaagcagttaccagaaaagtagcaaaatatctcaagaaaaatgaaattgggaagaaacg 1923
DB 244 TGATGAAGCAGTACCAGAAAGTAGCAAAATATCTCAAGAAAATGAAATTTGGGAAGAAACG 185
QY 1924 tgaactcaaggaagactcactatgtcagctaaagaaatctcaacaatgacaaacttca 1983
DB 184 TGAACCTCAAGGAAGACTCACATGTGTGCTAGCTAAAGAAATATCTTACAAATGGCAAACTTCA 125
QY 1984 ggatgatagtgagatgcttccaaaaaagctgttattgactgaatttagatcactcgtgtgat 2043
DB 124 GGATGATAGTGAGATGCTTCCAAAAAGAGCTGTTATTGACTGAATTTAGATCAGTGGTGAT 65
QY 2044 taaaaactctactccagaaaatcgcgtctggcataaataatgattatggtcactcaaaaaa 2103
DB 64 TAAAAACTCTACTTCCAGAAATCCGCTCTGGCATAAATGATGATTATGGTCAACTAAAAAA 5
QY 2104 ttcc 2107
DB 4 TTTC 1

RESULT 9
AAF44923
ID AAF44923 standard; cDNA; 396 BP.
XX
AC AAF44923;
XX
DT 28-MAR-2001 (first entry)
DE Human breast cancer related protein coding sequence SEQ ID NO: 79.
XX
KW Human; breast cancer; diagnosis; therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200078960-A2.
XX
PD 28-DEC-2000.
XX
XX 23-JUN-2000; 2000WO-US17536.
XX
XX 23-JUN-1999; 99US-0140903.
XX
XX 12-OCT-1999; 99US-0158980.
XX
XX (CORI-) CORIXA CORP.
XX
XX Yuqiu J, Mitcham JL;
XX
XX WPI; 2001-041426/05.
XX
XX New polynucleotides encoding breast tumor specific proteins, useful for
XX prevention, treatment and diagnosis of breast cancer -
XX
XX Claim 25; Page 146; 165pp; English.
XX
XX The present invention provides the coding sequences for a number of
XX breast cancer related proteins. These can be used in vaccinations
XX against, diagnosis of and treatment of cancer, particularly breast
XX cancer.
XX
XX SQ Sequence 396 BP; 124 A; 52 C; 45 G; 160 T; 15 other;

Query Match 8.1%; Score 358.4; DB 22; Length 396;
Best Local Similarity 93.1%; Pred. No. 1.3e-64;
Matches 353; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

QY 3978 tagatgcaagtctcttgycgatatagctattgagatatattacacttagtctgtggttga 4037
```


Db 1 tanatgcaagtnttngtgatatacgtatttgagataattacnctagctgtgcttga 60
QY 4038 ctgtttttttatgtttttgatgaatagaagttttaaattttgacaaagtcgaatttat 4097
Db 61 ctgtttttttatgtttttgatgaatanaagttttaaattttgacaaagtcgaatttat 120
QY 4098 tttttttttttgttatattttttctctccaatttaaccccaagatttcagatatctg 4157
Db 121 tttttttttgttgatattttttctccaatttaaccccaagatttcacatatntg 180
QY 4158 ctctattatataaactttatatatttttatatttgatctaccttgaaattgatgtatgt 4217
Db 181 ctctattatanaaactttatatatttttatatttgatctncttgaaattgatgtatgt 240
QY 4218 tttgaattatggaatgaggtgtttttttcccccatacagaattccagtcattgtaaac 4277
Db 241 tttgaattatggaatgaggtgtttttttcccccatacagaattccagtcattgtaaac 300
QY 4278 tgtttattgaaagaattatcttctctcattaaattacattgccaattagtaaaatca 4337
Db 301 tgtttattgaaanaattatcttctctcattaaattnccttgccaattagtaaaatca 360
QY 4338 attaacacatrmarmmmrrr 4356
Db 361 attaccataataaaaaaaa 379

RESULT 10
AAT24560
ID AAT24560 standard; cDNA to mRNA; 195 BP.
XX
AC AAT24560;
XX
DT 25-SEP-1996 (first entry)
XX
DE Human gene signature HUMGS06610.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX Homo sapiens.
XX
PN WO9514772-A1.
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP01916.
XX
PR 12-NOV-1993; 93JP-0355504.
XX
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
XX
PI Matsubara K, Okubo K;
XX
PP WPI; 1995-206931/27.
XX
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 1639; 2245pp; Japanese.
XX
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAT19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
SQ Sequence 195 BP; 80 A; 24 C; 31 G; 60 T; 0 other;

Query Match 4.4%; Score 193; DB 16; Length 195;
Best Local Similarity 100.0%; Pred. No. 1.5e-30;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2276 gatctttttgatacaatccttattttaaaggagaagataaactgagatttttaaaaga 2335
Db 1 gatctttttgatacaatccttattttaaaggagaagataaactgagatttttaaaaga 60
QY 2336 agccatggaaaaacttctctagtagcatctacttcaggcccaacaagggttatgaatata 2395
Db 61 agccatggaaaaacttctctagtagtagcatctacttcaggcccaacaagggttatgaatata 120
QY 2396 tagtgtatagaagcatttaagttacaatgtttttatgctcctaaattttattataaaatg 2455
Db 121 tagtgtatagaagcatttaagttacaatgtttttatgctcctaaattttattataaaatg 180
QY 2456 cacaaaaacttga 2468
Db 181 cacaaaaacttga 193

RESULT 11
AAK68762
ID AAK68762 standard; DNA; 6440 BP.
XX
AC AAK68762;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23574.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216847.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-02559678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

Disclosure; SEQ ID NO 23574; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patient's own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I) by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/hematopoietic-related diseases, especially cancers and cancer metastases of hematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/hematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAK82169


```
RESULT 14
AAT25563
ID AAT25563 standard; cDNA to mRNA; 121 BP.
XX AC
AC AAT25563;
XX XX
XX 15-NOV-1996 (first entry)
XX XX
XX Human gene signature HUMGS07740.
XX XX
XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX OS
XX Homo sapiens.
XX PN
XX WO9514772-A1.
XX PD
XX 01-JUN-1995.
XX PF
XX 11-NOV-1994; 94WO-JP01916.
XX PR
XX 12-NOV-1993; 93JP-0355504.
XX PA
XX (MATS/) MATSUBARA K.
XX PA (OKUB/) OKUBO K.
XX PI
XX Matsubara K, Okubo K;
XX XX
XX WPI; 1995-206931/27.
XX DR
XX
XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX PS
XX Claim 1; Page 1859; 2245pp; Japanese.
XX CC
XX A single-stranded DNA (or its complementary strand or the corresp.
XX double-stranded DNA) which comprises one of the 7837 "GS" sequences
XX given in AAT19001-T26837 and which is able to hybridise to part of
XX human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
XX sequences were obtained from 3'-directed cDNA libraries prepared
XX from various human tissues; synthesis of cDNA was initiated from the
XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
XX untranslated sequence is unique to a particular mRNA species, almost
XX all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
XX is constructed so as to reflect accurately the relative abundance of
XX different mRNAs in the particular tissue from which it was derived.
XX The appearance frequency of a given GS in a cDNA library can be
XX determined (esp. using primers and probes derived from the GS
XX sequences) as a means of diagnosing abnormal cell function or for
XX recognising different cell types.
XX SQ
XX Sequence 121 BP; 41 A; 25 C; 12 G; 43 T; 0 other;

Query Match 2.7%; Score 120.2; DB 16; Length 121;
Best Local Similarity 98.3%; Pred. No. 1.4e-15;
Matches 119; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4229 gatcagggtctctttttcccccatacaagatccagtcatttaacactgtttattgaa 4288
Db 1 gatcagggtctctttttcccccatacaagatccagtcatttaacactgtttattgaa 60

QY 4289 agaattatctcttcctattaaattaccttgccaattagtaaaaaatacaattaccatrm 4348
Db 61 agaattatctcttcctattaaattaccttgccaattagtaaaaaatacaattaccatrm 120

QY 4349 a 4349
Db 121 a 121
```

```
RESULT 15
AAK06689
ID AAK06689 standard; DNA; 442 BP.
XX XX
AC AAK06689;
XX XX
XX 05-NOV-2001 (first entry)
XX XX
XX Human brain expressed single exon probe SEQ ID NO: 6680.
XX DE
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX KW
XX OS
XX Homo sapiens.
XX XX
XX WO200157275-A2.
XX PN
XX 09-AUG-2001.
XX PD
XX
XX 30-JAN-2001; 2001WO-US00667.
XX PF
XX
XX 04-FEB-2000; 2000US-0180312.
XX PR
XX 26-MAY-2000; 2000US-0207456.
XX PR
XX 30-JUN-2000; 2000US-0608408.
XX PR
XX 03-AUG-2000; 2000US-0632366.
XX PR
XX 21-SEP-2000; 2000US-0234687.
XX PR
XX 27-SEP-2000; 2000US-0236359.
XX PR
XX 04-OCT-2000; 2000GB-0024263.
XX XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI
XX WPI; 2001-483446/52.
XX DR
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX PT
XX Example 4; SEQ ID NO: 6680; 650pp + Sequence Listing; English.
XX PS
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX CC
XX Sequence 442 BP; 153 A; 67 C; 80 G; 142 T; 0 other;

Query Match 2.7%; Score 119.4; DB 22; Length 442;
Best Local Similarity 99.2%; Pred. No. 2.9e-15;
Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2118 aaaaggtcacatattcctggagcagagaaaacttccacacatcattggaggatcagatctaa 2177
Db 264 aataggtcacatattcctggagcagagaaaacttccacacatcattggaggatcagatctaa 323

QY 2178 tagctcatcattgctgaagaataacacagactagaagagtgctgaaggcaggaatggagg 2237
Db 324 tagctcatcattgctgaagaataacacagactagaagagtgctgaaggcaggaatggagg 383

QY 2238 t 2238
Db 384 t 384

Search completed: August 15, 2002, 00:05:32
Job time: 9167 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 14, 2002, 23:56:50 ; Search time 58.32 Seconds
(without alignments)
1436.036 Million cell updates/sec

Title: US-09-837-602-2
Perfect score: 3899
Sequence: 1 MWKLLPAAGPAGGEPYRLLT.....KEESLADLFYRNPKRRR 754

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

| Database : | | | |
|--------------------|---|--|--|
| A_Geneseq_032802.* | | | |
| 1: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.* | | |
| 2: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.* | | |
| 3: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.* | | |
| 4: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.* | | |
| 5: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.* | | |
| 6: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.* | | |
| 7: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.* | | |
| 8: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.* | | |
| 9: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.* | | |
| 10: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.* | | |
| 11: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.* | | |
| 12: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.* | | |
| 13: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.* | | |
| 14: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.* | | |
| 15: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.* | | |
| 16: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.* | | |
| 17: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.* | | |
| 18: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.* | | |
| 19: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.* | | |
| 20: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.* | | |
| 21: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.* | | |
| 22: | /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.* | | |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Length | ID | Description |
|------------|-------|-------|--------|----|-----------------------------|
| 1 | 3893 | 99.8 | 754 | 21 | AAV51669 Human nibrin prote |
| 2 | 3885 | 99.6 | 754 | 21 | AAV32373 Nibrin, encoded by |
| 3 | 2046 | 52.5 | 399 | 22 | ABR92565 Human protein sequ |
| 4 | 473 | 12.1 | 811 | 22 | ABR62661 Drosophila melanog |
| 5 | 205 | 5.3 | 38 | 22 | AAW60530 Human bone expres |
| 6 | 205 | 5.3 | 38 | 22 | AAW73186 Human protein SEQ |
| 7 | 205 | 5.3 | 38 | 22 | AAW33396 Peptide #7433 enco |
| 8 | 171.5 | 4.4 | 990 | 22 | AAW78520 Human protein SEQ |
| 9 | 171 | 4.4 | 1812 | 22 | ABR58022 Drosophila melanog |
| 10 | 171 | 4.4 | 3257 | 22 | ABR67502 Drosophila melanog |
| 11 | 169.5 | 4.3 | 2918 | 22 | ABG27218 Novel human diagno |

| | | | | | |
|----|-------|-----|------|----|-----------------------------|
| 12 | 169 | 4.3 | 1128 | 19 | AAW40539 Mutant C-beta prot |
| 13 | 169 | 4.3 | 1128 | 21 | AAV84461 Amino acid sequenc |
| 14 | 168 | 4.3 | 808 | 22 | ABG05140 Novel human diagno |
| 15 | 168 | 4.3 | 832 | 22 | ABR93691 Human protein sequ |
| 16 | 167.5 | 4.3 | 931 | 22 | AAW79504 Human protein SEQ |
| 17 | 167.5 | 4.3 | 2400 | 22 | ABG20278 Novel human diagno |
| 18 | 167.5 | 4.3 | 2415 | 22 | ABG20279 Novel human diagno |
| 19 | 167 | 4.3 | 789 | 22 | AAW5460 Human protein sequ |
| 20 | 165 | 4.2 | 1411 | 17 | AAW02258 Nucleolar/endosoma |
| 21 | 165 | 4.2 | 1886 | 19 | AAW54241 Rattus norvegicus |
| 22 | 164.5 | 4.2 | 1392 | 20 | AAV06999 Restin protein seq |
| 23 | 164.5 | 4.2 | 2703 | 22 | ABR60074 Drosophila melanog |
| 24 | 163.5 | 4.2 | 2139 | 22 | ABR47278 PN7771. Homo sapi |
| 25 | 163.5 | 4.2 | 3111 | 22 | ABR60327 Drosophila melanog |
| 26 | 163.5 | 4.2 | 6815 | 22 | ABR66811 Drosophila melanog |
| 27 | 163 | 4.2 | 1164 | 19 | AAW40541 Mutant C-beta prot |
| 28 | 162.5 | 4.2 | 1427 | 12 | AAW10534 Human 160KD mediat |
| 29 | 161 | 4.1 | 1093 | 19 | AAW40540 Mutant C-beta prot |
| 30 | 161 | 4.1 | 1129 | 21 | AAW44462 Amino acid sequenc |
| 31 | 160 | 4.1 | 1099 | 19 | AAW40538 Mutant C-beta prot |
| 32 | 159.5 | 4.1 | 1833 | 22 | ABR71141 Drosophila melanog |
| 33 | 159 | 4.1 | 2096 | 21 | AAW41592 Human ORFX ORF1356 |
| 34 | 158.5 | 4.1 | 1664 | 20 | AAW9462 C. albicans alpha-1 |
| 35 | 158.5 | 4.1 | 2663 | 22 | AAW39097 Human polypeptide |
| 36 | 158 | 4.1 | 726 | 22 | ABR65327 Drosophila melanog |
| 37 | 158 | 4.1 | 1780 | 22 | AAW38681 Human polypeptide |
| 38 | 158 | 4.1 | 2897 | 22 | ABR58514 Drosophila melanog |
| 39 | 158 | 4.1 | 3263 | 22 | ABR67210 Drosophila melanog |
| 40 | 157.5 | 4.0 | 976 | 22 | AGG6581 Human SCP-1 mutain |
| 41 | 157.5 | 4.0 | 1552 | 22 | ABR71764 Drosophila melanog |
| 42 | 157 | 4.0 | 1164 | 17 | AAW85781 Group B Streptococ |
| 43 | 157 | 4.0 | 1164 | 19 | AAW40537 Group B streptococ |
| 44 | 157 | 4.0 | 1164 | 21 | AAW84459 Amino acid sequenc |
| 45 | 157 | 4.0 | 2665 | 22 | ABR28314 Human peptide #965 |

ALIGNMENTS

| | |
|----------|--|
| RESULT 1 | |
| AAV51669 | AAV51669 standard; Protein; 754 AA. |
| XX | AC |
| XX | AAV51669; |
| XX | 01-JUN-2000 (first entry) |
| XX | Human nibrin protein. |
| XX | |
| XX | Nibrin; human; DNA double strand break repair protein; diagnosis; |
| KW | therapy; Nijmegen Breakage Syndrome; gene therapy. |
| XX | Homo sapiens. |
| XX | DE19818680-Cl. |
| XX | 09-MAR-2000. |
| XX | |
| PF | 27-APR-1998; 98DE-1018680. |
| XX | |
| PR | 27-APR-1998; 98DE-1018680. |
| XX | |
| PA | (UYBE) UNIV BERLIN HUMBOLDT. |
| XX | |
| DR | WPI; 2000-196117/18. |
| XX | N-PSDB; AA289047. |
| PT | A DNA double strand break repair protein, Nibrin, and related DNA |
| PT | useful for diagnosis and therapy of Nijmegen Breakage Syndrome and |
| XX | other diseases influenced by DNA-double-strand break repair - |
| PS | Claim 1; Fig 2; 32pp; German. |
| XX | |

CC This invention describes a novel DNA double strand break repair protein,
CC Nibrin. Nibrin and DNA encoding it are useful for diagnosis and/or
CC therapy of diseases influenced by repair of DNA-double strand breaks,
CC in particular Nijmegen Breakage Syndrome. The product of the invention
CC has applications in gene therapy. This sequence represents the nibrin
CC protein described in the invention.

XX
SQ Sequence 754 AA;

Query Match 99.8%; Score 3893; DB 21; Length 754;
Best Local Similarity 99.8%; Pred. No. 9.5e-290;
Matches 752; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWKLLPAAGPAGGPPYRLTLTGVEYVVGKKNCAIENDQSIERNHVLITANFVSNTLSQT 60
DB 1 mwklilpaagppgpyrllltgveyyvgrkncailliendqsisrnhavltanfsvtnlsqt 60
QY 61 DEIPVLTKDNSKYGTFTVNEEKMONGFSRTLKSGDGTFTGFGSKFRIEYPLVACSSCL 120
DB 61 deipvltikdnskylgtftvneekmqngfsrtlksdgtitfgvfgskfrieypvacsscl 120
QY 121 DVSGKTALNOAILLOLGGFTVNNWTEECTHLVMVSVKVTIKTICALICGRPIVKPEYTFEF 180
DB 121 dvsgktalnailloolgftvnnwteecthlvmvsvkvtkticalicgrpivkpeytfef 180
QY 181 LKAVOSKRQPPQIESFYPPDLDEPSIGSKNVDLSGRQERKQTFKGTFTFLNAKQHKLSS 240
DB 181 lkaveskkrppqiesfyppdldepsigsknvdlsgrqerkqtfkgtftflnakqhkklss 240
QY 241 AVFEGGEARLTENEENHFFLAPGTCVVDTGITSQTLIPDCQKKWQSIQIMDLQRO 300
DB 241 avfeggearlteneeenhfflapgtcvtvdtgitsqtlipdcqkkwqsiqimdlqro 300
QY 301 GLRPIPEAETGLAVIENTTKNYCDQGHPSGLTKTTPGSLSGVSDKLMPSAPVNT 360
DB 301 glrpipeaetglavfenttknycdqghpsgltktttpgslsgvsvdekmpsapvnt 360
QY 361 TTYVADTESEQADTWLSEPKIEKVKMEQKFRMLSQDAPTVMKESCKTSNNNSMVSNT 420
DB 361 ttyvadteseqadtwlserpkiekvskmekfmlsqdaptvmkescktsnnnsmvsn 420
QY 421 LAKMRIPNYQLSPKPLSINKSKDRASQOQTNSIRNYFQSTPKKRDEENQEMSSCKS 480
DB 421 lakmrinyqlspkplsinkskdrasqoqtnsirnyfqpstkkredeeqemsscks 480
QY 481 ARIETCSLLEOTOPATPPLSKWKEQHLSENEPVDTSNNLFTDCLKSTVKNSASKSH 540
DB 481 arietcsllleotopatpplskwkeqhlsenepvdtssnnlftdclkstvknsasksh 540
QY 541 AAEKLSNRKREMDVVAIEDVLEQLFKDTKPELEIDVKQKQEDVNVKRPRMDIETN 600
DB 541 aaeeklsnrkremdvvaiedvleqlfkdtkpeleidvkqkqeedvnmkrprmdietn 600
QY 601 DTFSDEAVPSSKISQENEGCKRELKEDSLWSAKELSNNDKLQDDSEMLPKLLLTFR 660
DB 601 dtfsdeavpsskisqueenegckrelkedsllwsakelsnndklqddsemplklllfr 660
QY 661 SLVTKNFTSRNPSGINDYDGLKFKFKVYTPGACKLPHIIGSDLIHARKNTELE 720
DB 661 slvtnkftsrnpsgindydgllkfkfkvytpgacklphiiigsdliaharkntele 720
QY 721 EWLRCQMEVQNHAKESLADDLFRYPNPKRRR 754
DB 721 ewlrcqmevqnqhakesladdlfrypnplkrrr 754

RESULT 2
ID AAY32373
XX
AC AAY32373;
XX

DT 28-FEB-2000 (first entry)
XX Nibrin, encoded by Nijmegen breakage syndrome gene NBS1.
DE
XX
XX NBS1 gene; nibrin; Nijmegen breakage syndrome; diagnosis; human;
KW gene therapy; cancer; microcephaly; mental retardation;
XX primary ovarian failure.
XX
OS Homo sapiens.

XX Key Location/Qualifiers
FT Domain 24..100
FT /note= "forkhead associated domain"
FT 109..195
FT Domain
XX /note= "breast cancer carboxy-terminal domain"

XX W09955716-A1.
XX 04-NOV-1999.
XX 27-APR-1999; 99WO-US09036.
XX 27-APR-1998; 98US-0083269.
XX (VIRG-) VIRGINIA MASON RES CENT.
XX Concannon PJ, Vissinga CS, Cerosaletti KM, Varon R, Sperling K;
PI Reis A;
XX
XX WPI; 2000-062015/05.
DR N-PSDB; AAZ34997.

XX Novel gene useful for detecting mutations or polymorphisms, and
XX diagnosing certain pathological conditions in Nijmegen Breakage
XX syndrome patients.
XX
XX Claim 1; Fig 2A-B; 58pp; English.

XX This sequence represents nibrin, the protein encoded by the NBS1
XX gene (see AAZ34997) that is associated with the Nijmegen breakage
XX syndrome (NBS). The invention is based on the discovery that the
XX NBS1 gene contains mutations in all NBS patients. These mutations
XX include deletions and insertions that result in frameshift, as well
XX as point mutations. It is an object of the invention to detect a
XX mutation or polymorphism in NBS patients, and hence to diagnose a
XX predisposition to a pathological condition such as cancer,
XX microcephaly, mental retardation, and primary ovarian failure,
XX based on detection of a mutation in the NBS1 gene. It is also an
XX object of the invention to treat NBS by replacing the mutated gene
XX in a NBS patient by gene therapy. Recombinant vectors, genetically
XX engineered host cells, a method for producing nibrin polypeptide,
XX an antibody that specifically binds to the polypeptide, and a
XX method for diagnosing NBS by detection decreased expression of the
XX gene or by detecting a mutation in the gene are claimed.

XX Sequence 754 AA;

Query Match 99.6%; Score 3885; DB 21; Length 754;
Best Local Similarity 99.6%; Pred. No. 3.9e-289;
Matches 751; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MWKLLPAAGPAGGPPYRLTLTGVEYVVGKKNCAIENDQSIERNHVLITANFVSNTLSQT 60
DB 1 mwklilpaagppgpyrllltgveyyvgrkncailliendqsisrnhavltanfsvtnlsqt 60
QY 61 DEIPVLTKDNSKYGTFTVNEEKMONGFSRTLKSGDGTFTGFGSKFRIEYPLVACSSCL 120
DB 61 deipvltikdnskylgtftvneekmqngfsrtlksdgtitfgvfgskfrieypvacsscl 120
QY 121 DVSGKTALNOAILLOLGGFTVNNWTEECTHLVMVSVKVTIKTICALICGRPIVKPEYTFEF 180
DB 121 dvsgktalnailloolgftvnnwteecthlvmvsvkvtkticalicgrpivkpeytfef 180

QY 181 LKAVQSKKOPQIESFYPLDEPSIGSKNVDLSGRQERKQIFKGTFFIFLNAKOHKKLSS 240
Db 181 lkaveskkppqiesfyppldepsigsknvdlsgrqerqifkgktfflnakqhkkllss 240
QY 241 AVVFGGGEARLITEENEEHNFPLAGTCVVDGTGINSQTLIPDCOKKWTQSIQMDMLQK 300
Db 241 avvfgggearliteeneeehnfflapgtcvvdgtginsqtlipdcqckkwigsimdmqlrq 300
QY 301 GLRPIPEAEIGLAVIFMTTKNYCDPQGHPSGTGLKTTTPGPSLSQGVSDVKLMPSPVNT 360
Db 301 glrpipeaeiglavifmttknycdpghpstgltkttpgpslsqgvsdvdklmpsapvnt 360
QY 361 TTYVADTESQADTWDLSEKPKIEKYSKMEQKFRMLSODAPTKECKTSSNNNSMVSNT 420
Db 361 ttyvadteseqadtwdlserpkeikyskmeqkfrmlsqdaptkvescktsnnsmvsnt 420
QY 421 LAKMRIPNQLSTKIPSLPSINKSKDRASQOQOTNSIRNYQPSYTKKRRERDEENQESCKS 480
Db 421 lakmripnyqlsptklpslnkskdrasqqqgtqnsirnyqpsytkkrrerdeengemscks 480
QY 481 ARTECSLLEQTOPATPSLWKNKEOHLSENPVDTNSDNNLFTDIDLKSIIVKNSASKSH 540
Db 481 arletsclleqtpatpslwnkedhlsepevdtnsdnnlftdtdlksivknsasksh 540
QY 541 AAELRSNKKREDDVAIEDVLEQLFKDTKPELEIDVKYQKEEDVNVKRRPRMDIETN 600
Db 541 aaeklrnskkremddvaiedevleqlfdtkpeleidvkvqkeedvnrkrprmdietn 600
QY 601 DTFSDEAVPESKISQENEIGKREKLEKEDSLWSAKEISNNNDKLQDDSEMLPKKLLITEFR 660
Db 601 dtfsdeavpesskisqeneigkreikedslwsakeisnndklqddsemplpklllitefr 660
QY 661 SLVIKNTSRNPSSINDDYCOLKNFKFKKVTYPGAGKLPHIIGGSLIAHHARKNTELE 720
Db 661 slvikntsrnpssinddyqolkfnfkfkvtypgagklphiiggslliahharkntele 720
QY 721 EWLRQMEVQNHAKESLADDLFRYNPYLKRKR 754
Db 721 ewlrqemevnqhakeesladdlfrynpylkrkr 754
RESULT 3
AAB92565
ID AAB92565 standard; Protein; 399 AA.
XX AAB92565;
AC AAB92565;
XX 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:10768.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000BP-0116126.
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX

DR WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 8; SEQ ID 10768; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX Sequence 399 AA;
SQ
Query Match 52.5%; Score 2046; DB 22; Length 399;
Best Local Similarity 98.7%; Pred. No. 1.7e-148;
Matches 393; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 MWKLLPAAGPAGGEPYRLLTGVEYVVGKNCALIIENDQSISRNVHVLTFANFSVTNLSOT 60
Db 1 mwkllpaagpaggepyrlltgvvyvgrkncallieendqslsrnhavltanfsvtnlsqt 60
QY 61 DEIPVLTLDKNSKYGTFTVNEEKQNGFSRTLKSGDGTTFGVFGSKFRIEYPLVACSSCL 120
Db 61 deipvltldknskylgtftvneekmgngfsrtlksgdgttfvgfskfrieyplvacsscl 120
QY 121 DVSGKTALNOAILQLGGFTVNNWTEECTHLVMVSVKVTIKTICALICGRPIVRPEYTFEF 180
Db 121 dvsgktalnqailqlggftvnnwteecthlvmvsvkvtkikticalicgrplvkrpeytfef 180
QY 181 LKAVQSKKOPQIESFYPLDEPSIGSKNVDLSGRQERKQIFKGTFFIFLNAKOHKKLSS 240
Db 181 lkavqskkoppqiesfyppldepsigsknvdlsgrqerqifkgktfflnakqhkkllss 240
QY 241 AVVFGGGEARLITEENEEHNFPLAGTCVVDGTGINSQTLIPDCOKKWTQSIQMDMLQK 300
Db 241 avvfgggearliteeneeehnfflapgtcvvdgtginsqtlipdcqckkwigsimdmqlrq 300
QY 301 GLRPIPEAEIGLAVIFMTTKNYCDPQGHPSGTGLKTTTPGPSLSQGVSDVKLMPSPVNT 360
Db 301 glrpipeaeiglavifmttknycdpghpstgltkttpgpslsqgvsdvdklmpsapvnt 360
QY 361 TTYVADTESQADTWDLSEKPKIEKYSKMEQKFRMLSODAPTKECKTSSNNNSMVSNT 420
Db 361 ttyvadteseqadtwdlserpkeikyskmeqkfrmlsqdaptkvescktsnnsmvsnt 420
RESULT 4
AAB92661
ID AAB92661 standard; Protein; 811 AA.
XX AAB92661;
AC AAB92661;

XX
PS Example 4; SEQ ID NO: 32635; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 38 AA;

Query Match 5.3%; Score 205; DB 22; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.5e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 691 VTPVGAGKLPHTIGGSDLIHARKNTELEEWLRQEME 728
Db 1 vtypgagkplhliggsdliaharknteleeewlrqeme 38
|||||

RESULT 6
AAM73186
ID AAM73186 standard; Protein; 38 AA.
XX
AC AAM73186;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 33492.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX

PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR -30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -

PS Example 4; SEQ ID NO: 33492; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
SQ Sequence 38 AA;

Query Match 5.3%; Score 205; DB 22; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.5e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 691 VTPVGAGKLPHTIGGSDLIHARKNTELEEWLRQEME 728
Db 1 vtypgagkplhliggsdliaharknteleeewlrqeme 38
|||||

RESULT 7
AAM33396
ID AAM33396 standard; Protein; 38 AA.
XX
AC AAM33396;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #7433 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488997/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -

PS Claim 27; SEQ ID No 33665; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs;
CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
SQ Sequence 38 AA;
Query Match 5.3%; Score 205; DB 22; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.5e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 691 VTPVGAGKLPHTIGGSDLIHARKNTELEEWLRQEME 728
Db 1 vtypgagkplhliggsdliaharknteleeewlrqeme 38
|||||

RESULT 8
AAM78520
ID AAM78520 standard; Protein; 990 AA.
XX
AC AAM78520;

[illegible]

Db 1288 risren-----rdlkdeivrlndsilqateqrrraeenalqgkacgseimqk-----kqhl 1338

QY 575 EIDVK-VQOEEDVNVKRPMDIETNDT-----FDEAVPESKISQENE 619

Db 1339 eileikvmqgreednarhkqsie-eaaktiqdnkneierlkaefqea---krrweyene 1394

QY 620 IGKKRELKEDSLWSAK-----EISNNDKLQDDSEMLPKRLLLT 657

Db 1395 lskvrnnydeeiislknqfeteinicktthqitmqkeedtsqyraqidnltrenselse 1454

QY 658 EFRSLVIKNS---TSRNPISGINDYQQLKNFKKVVTPGAGKLPHIIGGSDLIHAHAR 714

Db 1455 eikrl--kntlqttenlrirvedliq-----qkat-----gsev-----sq 1489

QY 715 KNTLEEWLRQ--EMEVQNHAKESLAD 741

Db 1490 rkqgleveirvqtgmrtesvrykqslidd 1518

RESULT 12

AAW40539

ID AAW40539 standard; Protein; 1128 AA.

XX

AC AAW40539;

XX

DT 28-AUG-1998 (first entry)

XX

DE Mutant C-beta protein nv34qp.

XX

KW C-beta protein; beta antigen; vaccine; group B Streptococcus.

XX

OS Streptococcus sp. strain A909.

XX

XX Synthetic.

PN WO9809648-A1.

XX

PD 12-MAR-1998.

XX

PF 05-SEP-1997; 97WO-US15319.

XX

PR 06-SEP-1996; 96US-0024707.

XX

PA (NAVA-) NORTH AMERICAN VACCINE INC.

XX

PI Blake MS, Tai JY;

XX

XX

DR WPI; 1998-193324/17.

XX

DR N-PSDB; AAV11346.

XX

PT Group B streptococcal Cbeta proteins - having amino acid

PT substitutions to reduce immunoglobulin A binding while retaining

PT antigenicity, for use in vaccines

XX

PS Disclosure; Fig 7A-C; 59pp; English.

XX

CC Group B Streptococcus C-beta mutant protein nv34qp has only 10% of

CC the IgA binding activity of the native C-beta protein (see AAW40537).

CC It is encoded by mutated C-beta protein DNA (see AAV11346) in which

CC the Gln-167 and Gln-174 codons of the mature C-beta protein coding

CC region are replaced by pro codons. These residues are in the

CC IgA binding domain of the C-beta protein. The IgA binding ability

CC of C-beta appears to require dimerisation of the protein. The

CC invention relates to mutant C-beta proteins that have a reduced or

CC eliminated ability to bind human IgA but which retain the

CC antigenicity of the wild-type protein. Such proteins can be used

CC in vaccines for use against group B streptococci. Claimed mutant

CC proteins have mutations in the IgA binding region of C-beta.

XX

XX Sequence 1128 AA;

SQ

Best Local Similarity 19.9%; Pred. No. 0.00072;

Matches 142; Conservative 125; Mismatches 238; Indels 208; Gaps 38;

QY 192 QIESFYPPLDPEPSIGSKNVDSLGR-----QERKQIFKCTFFFLNAKHKKLSSA 241

Db 95 qkneftktdetndsdallelenqfnetnrlhlhikqheevkdk-----kakqgtlikqs 149

QY 242 VVEGGGEARLITBENEHEHFFLAPGTCVVVD-TGITNSOTLIIPDCOKKWTQSIQIMDLQK 300

Db 150 -----dtkvdisnidckelnhqkspvekmaepkgtined---kdsmlkkiedlrkqaqa 200

QY 301 GLRPIPEAEI--GLAVIFMTTKYCDPQHPSTGLTKTPPGPSLSQGVSVDEKLMFSAPV 358

Db 201 dkkedaevkvreelglfsstkgagldqeihvkkets-----seentqkvdehly-----a 251

QY 359 NTTTYVADTSEQADTWLSERPKEIKVSKME--QK-----FRLMSQDAP 401

Db 252 nslqnlaklsleeldkattneqatqvkqnfienaqklkemqliketnrvlykamsesle 311

QY 402 TVKESCKTSSNNN--SMVSNLTAKMRIPNYQLSPTK-LPSINKSKDRASO--QOOTNSIR 456

Db 312 qvekelkhseanledlvakskeivreyegklnqsknlpelqleeeahsklkqvvedfr 371

QY 457 NYF----QPSTKKR-ERD-----EENQEMSSCKSARIETSCSLLEQTOPTATPSLWKNKEQH 507

Db 372 kkfktseqvtppkkrvkrdlaanenngq-----kilelts-----penitvyegedvk 418

QY 508 LSENEPVDNSDNNLFTD-TDLKSIYKNSASKSHAAEKLRSKKNKREDDVAIEDVLEQL 566

Db 419 ft-----vtaksdsktldfisdL--ltkynpsvs---drisnkyktntdnkhaieftiknl 469

QY 567 -----FRDTRK---PELEIDVKVQOEEDVNVKRPMDIETNDTFSDEAVPES 612

Db 470 klnesqvtlikakddsgnvvektftitvqkkeek-qvpktp-----eqkdskteekvpqep 524

QY 613 KISQENEI-----GKKRELK-----EDSLW-SAKE-----IS 638

Db 525 ksndknqlqeliksaqqeleklekaikelmegpeipnpeyigkswesqkepiqeaif 584

QY 639 NNDKLQDDSS-----EMLPKKLLL-----TEFRSLVI 664

Db 585 sfkkliigdsesskytyehyfnkyksdfmnyqlhaqmemltrkvvqymnkypdnaelkk-if 643

QY 665 KNSTSRNPISGINDYQOLKN-----FKKF-----KKVTYPGAGKLPHI 702

Db 644 esdmkrtk---ednygslendalkgyfekyfltpfnkikivddldkxkveqdpapape- 699

QY 703 IGGSDLIHAHARKNTELEEWLRQEMEVQNHQ--AKEESLADDLFRYNPYLKRR 753

Db 700 --nsemdqakekakiaavskymskvldgvhqlqkknhsxivdflfkeleakqg 750

RESULT 13

AAW84461

ID AAW84461 standard; Protein; 1128 AA.

XX

AC AAY84461;

XX

DT 25-JUL-2000 (first entry)

XX

DE Amino acid sequence of a C-beta protein.

XX

KW C-beta protein; bactericidal; gram positive bacteria; vaccine;

KW Immune response.

XX

OS Streptococcus agalactiae.

XX

PN WO200015760-A1.

XX

PD 23-MAR-2000.

XX

PF 17-SEP-1999; 99WO-US21643.

XX


```
XX SQ Sequence 808 AA;
Query Match
Best Local Similarity 4.3%; Score 168; DB 22; Length 808;
Matches 125; Conservative 112; Mismatches 204; Indels 148; Gaps 32;
QY 252 ITEEN-----EEHN-----FFLAPGTCVVDGTG-TNSQTLIPDCQKWKIQSIMDMLQK 300
Db 109 lteenfrlqaehdqakelfllrktleemrlretqkql--nardeisikllmqlgsk 166
QY 301 GL-----RPIPAE--IGLAVIFMTKN-----YCDPQGHPTGLTKTTPGPSL 342
Db 167 gipskslednertirmaeaeqsvshvildqgkekenihlreelhrsqglq---pepak 223
QY 343 SOGVSDEKLPAPNTTYYADTDESE-----QADTWDLSEPRKEIKVSKMEQKFR 394
Db 224 tkalgtviemkdkiaslernirdiedelqmkangvlnedreeikqlevykhskfm 283
QY 395 -----MLSQDAPTAKES-----CKTSSNNMSVSNPLAKMRIPNYQLSPTKLPISINK 441
Db 284 ktidqkql-sktesellalqtkletlsgnqsdckqhievk-----eslta 331
QY 442 SKDRAS-QQQQNSIR-----NYFPSTKK-RERDEEN-----QEMSSCKSAR 482
Db 332 keqraailqtevdaillrleekesfinkktkqlqlditeekgtlageirdmkmlevkerk 391
QY 483 IETCSLLEOTQATPAPSLWKNEQHL-----ENEPVDYNSDNNLFTDTDLKSIVKNSASK 538
Db 392 lnvlqklienlqeql-----rkdkgkltlnkdrvkslqtdssn---tdtalatleeealsek 444
QY 539 SHAAKLRNSKREKMDVAIEDVLEOLFQKDKPELEIDVYKQKQEDVNVKRPRMDI- 597
Db 445 erlierlk--eqrerdd-----rerleesfrkenkdilkeknalqaeitekesslidk 498
QY 598 ETNDTFSDEAVPESSKI-SQENEIGKKRELKEDSLWSAKEISNNDKLODDSEMLPKLLL 656
Db 499 ehassiasaglkrdsklslelaieqkke--ecsklea-qlkkahnieddsrmnp----- 550
QY 657 TEFERSLIVKNSRNPNGINDYG-----OLKNFK--KFKKVTPYGAGKLP 700
Db 551 -efadqil--kqldkeasyrdecgkaqaevdrllleilkevenekndkdkkiaelesltir 607
QY 701 HIIGSDLIAH-----HARNTLEEWLRQEMEVONHAKESLADD 742
Db 608 hmkdgnkvvanlkhnqglekknqnaqlleevrr-----redsmadn 647
RESULT 15
AAB93691
ID AAB93691 standard; Protein; 832 AA.
XX
AC AAB93691;
XX
XX 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:13253.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
```

```
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 13253; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 832 AA;
Query Match 4.3%; Score 168; DB 22; Length 832;
Best Local Similarity 21.1%; Pred. No. 0.00055;
Matches 133; Conservative 100; Mismatches 250; Indels 148; Gaps 27;
QY 185 OSKKQPPQIESFY---PPLDEPSIGSKNYVDLSGRQERKQIFKGTIFLNAKQHKLLSSA 241
Db 142 eakktgptifspfcsmpp1-fptvgkdkvn-----nlladpenivtyknreanmdfs 193
QY 242 VVFGGGEARLITEENEENHFFLAPGTCVVDGTGTTNSQTLIPDCQKWKIQSIMDLQ--- 298
Db 194 svf--spsfqipevceeneyfcp---vnsngklsdlsiltnepvk-----rdsldashy 242
QY 299 ---RQGLRPIPEAEIGLAVIFMTKNKYCD-----PQGHPTGLTKTTPGPSLSQ 345
Db 243 sgvsqnir--skaqi-lallkssesssceeInsemtehfpkgqpggsklatpkylq 299
QY 346 VSDEK-----LMPAPVNTTYYA-DTDESE-QADTWDL 377
Db 300 eecaemstenlyyqhgsentmrnkswamy1ssgssphssthsvdgnndtekrpkageddv 359
QY 378 SERPKKIKVSKMEQKFRMLSQDAPTVKESCTSSNNMSVSNPLAKMRIPNYQLSPTKLP 437
Db 360 nsnlklsiqkiiqfvetyaeerk--kynvdqsvgndpswqevkleipsf----- 409
QY 438 SINKSDRASQQQQTNSIRNYFQPTKKRERDEE---NQEMSSC-----KSARIETS 486
Db 410 -----nessslqvtcssaendgilssedlqdnkkipfnqndkgcivlikenaqevnt 464
QY 487 CSLLEOTQATPAPSLWKNEQHLSENEPVDYNSDNNLFTDTDLKSIVKNSASKHAAEKL 546
Db 465 cgtleke-----yeqssslpelkhlqiesnnrsrisddtdmises----- 506
```

```
QY 547 SNKKREMDVAIEDEVLEQLFKDTKPELETDVVKVQOEEDVNVKRPMDIETNDTSDE 606
Db :||: : : | | | | | : : : : : : : : : : : : : : : : : : : : : : : :
507 -----kmdnesl-nsiheslsnvtqpflevtfnl-----nfetsdt--ee 544
QY 607 AVPESSKISOENEIGKKRELKEDSLWSAKETSNNDKLOD-DSEMLPKKLLLTFRSLVIK 665
Db ||:||||:| | | | | : : : | | | | | : : : | | | | | : : : | | | | |
545 esqesnkisdqseswkdllvndgnsfcfkrsentnceeeieghp---fltsvdl-dk 600
QY 666 NSTSRNPS-GINDYGGOLKNFKK------FKKVTPGAGKLPHIIGGSDLIAHHARKNT 717
Db | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
601 ndehvlpstssdsvqlntnqnhyecialdksnthinslsfypigkhhlskdteahi 660
QY 718 ELEEWLRQEMEVOHQHAKESLADDLFRYP 748
Db | | : | | : | | : | | : | | : | | : | | : | | : | | :
661 sepedlgkirspppdhvevetaregkqwynp 691
```

Search completed: August 15, 2002, 01:40:35
Job time: 6225 sec

A;Molecule type: DNA
A;status: preliminary; translated from GB/EMBL/DDBB

34418
hypothetical protein F12F3.3 - *Caenorhabditis elegans*

F34418
hypothetical protein F12F3.3 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34418
R:Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of *C. elegans* cosmid F12F3.
A:Reference number: Z21521
A:Accession: T34418
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3488 <FUL>
A:Cross-references: PIDN: AAC25885.1; GSPDB: GN00023; CESP: F12F3.3
A:Experimental source: strain Bristol N2; clone F12F3
C:Genetics:
A:Gene: CESP:F12F3.3
A:Map position: 5
A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

| | | | | |
|-----------------------|-------------------|---|-------------|--------------|
| Query Match | 4.8% | Score 185.5; | DB 2; | Length 3488; |
| Best Local Similarity | 19.9% | Pred. No. 0.14; | | |
| Matches 132; | Conservative 114; | Mismatches 237; | Indels 181; | Gaps 31; |
| QY | 166 | ICGRPIVKPEYTFELKAVQSKK---QPPQTESYPPPLDEFESIGSKNVNDSLGRQBRKQIF | 222 | |
| DB | 466 | VTGRPL-----EAKPVEDKDAQSPSSKESPPTD---GKK-----KKQIP | 505 | |
| QY | 223 | KGKTFIFL-----NAKQHKLLSAAVVGGEAR-----LI | 252 | |
| DB | 506 | KA---LPIPDEISRRFGDPSTMHSETNITTTIRREGSADAKTPIVEPLSASVMKVETL | 562 | |
| QY | 253 | TEENDEEHFFLAPGTCVVDITGINSOTLIPDCQKKWIIQSIMDLQROGLRAPPEAETGL | 312 | |
| DB | 563 | VESAKERAFFSK-----RRET---PDKSR-----KKEGLPPAKKSE--- | 598 | |
| QY | 313 | AVIFMTTKNYCDPOGHPSTGLKTTTPGSLSQGVSVDE---KLMPASAP-----VNTTTY | 363 | |
| DB | 599 | -----KKDEVTAEQSTEALIESKKKEVDESKISEQOPSPDKNSEVVGVPKE | 645 | |
| QY | 364 | VADTESEQADWDLSERPKEIKVSKMEQKF-RMLSODAPTAKESCCTSSNNNSMVSNTLA | 422 | |
| DB | 646 | AAGPETKK-DVSEIEEVPVKTTIKKTEKSDSSISQSKSNVLKPADDDKSKSDVDVTKSKK | 704 | |
| QY | 423 | KMRIPNYQLSPTKLPSINKSKDRASQOQOTNSIRNYFQPTKKRDERDEQEMSSCKSAR | 482 | |
| DB | 705 | TTE-----DQTKVATDSKLEKAADTTKQIET-ETVVDKSKKKVLKKTEKSDSFISQK | 757 | |
| QY | 483 | IETSCSLLQEQTPAPSLWNKKEOHLSNEPVDNSNNLTDTDDLKSIIVNKSASKSHAA | 542 | |
| DB | 758 | SETP-PVVEPTKPAESEAQKTAENYNAKKQ--KEVDNLRKEAEV-----AAKKTAD | 806 | |
| QY | 543 | EKLK-----SNKKREMD-DVAIEDVLEQLFKDTPKELEIDVKKVQKQED | 586 | |
| DB | 807 | EKLIEAEANIKKTAEVEAAKKQEKDEQLKETEVWSKSSAAEKLKLEQAQIKKAAEA | 866 | |
| QY | 587 | VNVKRRPMDJETND-----TFSDAEVPSSKISOEN-EIGKKKRELK----- | 627 | |
| DB | 867 | DAVKQKQELN-EKKKLEAAKSAADKLKLEESAASKSVSESVKFGEEKTKRAGEKTV | 925 | |
| QY | 628 | --EDSLWSAKEISNND-KLQDSEMLPKLLLTLEFRSLVIKNSTSRNPSGIND-----DYG | 680 | |
| DB | 926 | QVESEPTSKKTIIDTKVGATEPADETPKK-----KIIKKTEKSDSSISQKSATDSE | 977 | |
| QY | 581 | QLKNFKKFKKVTYPGACKLPHIIIGSDLIHAHKNKNTLEEWLRQEMEV---QNOHAKEE | 737 | |
| DB | 978 | KVSKQKEQDEPTKPAVSETQMVTE-----ADKSKKQKETDEKLDLDAETAAKTQEADEK | 1032 | |
| QY | 738 | SLAD | 741 | |
| DB | 1033 | SKLD | 1036 | |
| RESULT 6 | | | | |

RESULT 6

S67593
transport protein US01 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein D252; protein YDL058w
C:Species: Saccharomyces cerevisiae
C:Date: 12-Jul-1996 #sequence.revision 12-Jul-1996 #text_change 21-Jul-2000
C:Accession: S67593; A38455; S30782
R:Blöcker, H.; Brandt, P.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67587
A:Accession: S67593
A:Molecule type: DNA
A:Residues: 1-1790 <BL>
A:Cross-references: EMBL:Z74106; NID:g1431058; PID:e353003; PID:g1431059; MIPS:YDL058w
A:Experimental source: strain S288C
R:Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamasaki, M.
J. Cell Biol. 113, 245-260, 1991
A:Title: A cytoskeleton-related gene, US01, is required for intracellular protein tra
A:Reference number: A38455; MUID:91185402
A:Accession: A38455
A:Molecule type: DNA
A:Residues: 1-389, 'TA', 392-724, 'S', 726-1790 <NAK>
A:Cross-references: GB:X54378; NID:g4777; PIDN:CAA38953.1; PID:g4778
A:Note: The authors translated the codon ACT for residue 768 as Ile
R:Hottel, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.
submitted to the EMBL Data Library, February 1993
A:Description: An integrin analogue in Saccharomyces cerevisiae.
A:Reference number: S30782
A:Accession: S30782
A:Molecule type: DNA
A:Residues: 71-846, 'E', 848-923, 'K', 925-1252, 'I', 1254-1318, 'V', 1320-1460, 'S', 1462-1580
A:Cross-references: EMBL:L03188

```

Query Match      4.6%; Score 179.5; DB 2; Length 1790;
Best Local Similarity 23.1%; Pred. No. 0.12;
Matches 100; Conservative 77; Mismatches 168; Indels 87; Gaps 19;

Qy 378 SERPKEIKVGMEOKFMRLSQDAPTVKESCKTSSNNNSVMTLAKMRIPNYOLSPTKLP 437
Db 988 SKNESITNLNQLKIDMSQERKNFOERGSTEKNIEQLKKTIS-----DLEQTEE 1040
Qy 438 SINK---SKDRASQO-----QQTNSIRNYFQPTKKRRERDEENQEMSWCKSARIE 484
Db 1041 IISKSDSSKDEYESQISLLKEKLETATTANDENVNKISELTKTRELEAEALAYKNLKNE 1100
Qy 485 TSCSLLBQTOPAPSLWKNKEQHLS-----ENEPVDNTSDNNLFTDIDLKSIVK---NS 535
Db 1101 LE-TKLTSETKALKEV-KENEEHLKEBKIOLEKBEATKQOOLNSLR-ANLESKEHEHD 1157
Qy 536 ASKSHAAEKILSRNKKRMDD--VAIEDVL-----EQLFKDTKPELEIDVVKOK--QEED 586
Db 1158 AAQLKYYEQIANKERYNEEISQLNDEITSTQOENESIKKKNDELGEVYKAMKSTSEQ 1217
Qy 587 VNVKRP-----RMDIETNDFSDAVPESKISQENIEIGKKRELKEDSLWASK 635
Db 1218 SNLAKSIDALNLQIKELKKKNETNEA---SLLESITK-SVESETVRIKELQDECNPFKEK 1272
Qy 636 EISN-NDKLQDSDMLPKLLLTFRSLVITKNSTSRNPNGINDDYGOLKNFKPKKVTYP 694
Db 1273 EVSLEDKLR-ASEDKNSKYLELQEKSEKIKLEELDAKTTTELKLTQLEKITNLSRAKESKES 1331
Qy 695 GAGKLPHIGGSOLIAHARKNTELE-EWLRQEMEVQONHAKEE----- 737
Db 1332 ELSLKK-----TSSEERKNAPLEKIKLNETOIKNOAFEPKRLKLLNGSSTITOEY 1384

```

| | | | | |
|--|---|--|----------|-----|
| Qy | 738 | ---- | SLADDLFR | 745 |
| | : | : | : : | : |
| Db | 1385 | EKINTLEDELIR | 1396 | |
| | | | | |
| RESULT | 7 | | | |
| T27522 | hypothetical protein ZC373.4 - Caenorhabditis elegans | | | |
| C:Species: | Caenorhabditis elegans | | | |
| C:Date: | 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000 | | | |
| C:Accession: | T27522 | | | |
| R:Kershaw, J. | submitted to the EMBL Data Library, April 1995 | | | |
| A:Reference number: | Z20382 | | | |
| A:Accession: | T27522 | | | |
| A>Status: | preliminary; translated from GB/EMBL/DDBJ | | | |
| A:Molecule type: | DNA | | | |
| A:Residues: | 1-1211 <WIL> | | | |
| A:Cross-references: | EMBL:Z49131; PIDN:CAA88976.1; GSPDB:GN00028; CESP:ZC373 | | | |
| A:Experimental source: | clone ZC373 | | | |
| C:Genetics: | | | | |
| A:Gene: | CESP:ZC373.4 | | | |
| A:Map position: | X | | | |
| A:Introns: | 20/1; 53/2; 80/3; 122/2; 169/3; 198/3; 236/2; 286/2; 337/3; 381/ | | | |
| | | | | |
| Query Match 4.5%; Score 177; DB 2; Length 1211; | | | | |
| Best Local Similarity 20.7%; Pred. No. 0.098; | | | | |
| Matches 171; Conservative 137; Mismatches 317; Indels 200; Gaps | | | | |
| Qy | 3 | KLLPAAGPAGGEP-----YRLTGVEYYVGR-----KNCAILLIENDQSISRNHAVLTANF | 52 | |
| | : | : : | : : | : : |
| Db | 455 | KLVTETDGDGLRPPPELMTKRASTGESLLAAVKVLKNLEV-VENGNSPKRK----- | 505 | |
| Qy | 53 | SVTNLSDTDIEIPVLTLKDNSKYGTFFNEKMONGSRITLKSGDGTFGVFGSKFRIEYEP | 112 | |
| | : | : : | : : | : : |
| Db | 506 | SLCSVKEEKIEP--AKKOTEGYKTL---ETTNG----- | 534 | |
| Qy | 113 | LVA CSSCLDYSGKTA-LNQAILQLQGFTVNNTTEECTHLVMVSVKVTIKTTICALICGRPI | 171 | |
| | : | : : | : : | : : |
| Db | 535 | -----DIRKSAPVKQDTL-----TCODNSKVTPQLLETAPNGELIRRSASAKEBS | 581 | |
| Qy | 172 | VKPEYFFEFLKAVOSKKQPQIESPYPLDEPSIGSKNNVDLSGRQERKQIFKGFTFIPLN | 231 | |
| | : | : : | : : | : : |
| Db | 582 | VKPEEVIR-KSSASVKETKTVELSVVPVKLEKLA--IDEGEKLVKK----- | 627 | |
| Qy | 232 | AKQHKLSSAVVFGGEARLITENEENHNFPLA PCTCVVDGTGINTSOTLPIDCOKWKIQ | 291 | |
| | : | : : | : : | : : |
| Db | 628 | ----KKTTTTATBIEKPKSIVIESKEE-----LVERPKLLINGTLVSDALEK--RR | 672 | |
| Qy | 292 | SI-----MDMLQRGLRPIPEAEIGLAVIFMTTKNYCDPGHPSTGLKTTTP-----G | 339 | |
| | : | : : | : : | : : |
| Db | 673 | SLGSKRASEFDVGLRPLPVWDSPKIVLTDKKEPKESKTKPAEPLKIVPETKKPARQG | 732 | |
| Qy | 340 | PSLSQGVSVDEKLMPASNVT---TYVADTSEQADTWDLSERPKEIKVSKMEQKFRLM | 396 | |
| | : | : : | : : | : : |
| Db | 733 | ALRRDKSLDD-ILNAAPSPTLLSKSTVTSVKHSVET--IHPSGEIIKSEKVLTKKSL | 789 | |
| Qy | 397 | SQD-----APTVKESCITYSSNNNSWSNTLIAMRIP-----NYQLSP | 433 | |
| | : | : : | : : | : : |
| Db | 790 | KKDSIDRLFPVKIIEDKASLPKLDVVVTFESKPVKYSKITVLK-IPE DANLIKDEKPP | 848 | |
| Qy | 434 | TKLPSINKSDRASQQOQTNSIRNYFPQSTKKRER--DEENQMSCKSARIETSCSILE | 491 | |
| | : | : : | : : | : : |
| Db | 849 | TKI--VMISEKAQTTVOTTTSVARTKSSKEKVHEHTADGKSVESSQKNKSORDD-----VK | 903 | |
| Qy | 492 | QTQPATPSLWNKEQHLSENPEVDYTNDSNNLTDTDLSKISVKNASKASHAEKLRSNKKR | 551 | |
| | : | : : | : : | : : |
| Db | 904 | VSQVTT-----KKEEDSTOPAP-----TLTVKKNVVKQTA EKSTSEKEHKHTTENK | 948 | |
| Qy | 552 | EMODVAIDEVLEQLFKD-TKPELEIDVKV--OKQEEDVNVNRKR--RMDIETNDITFS | 605 | |
| | : | : : | : : | : : |
| Db | 949 | EOERIIRKSSSTLKD--KDAOVAOSLOOSIKVNGEKKHESKVICTKVINKLKLGKETIN- | 1005 | |

```

QY   606 BAVPESS-----KISQEN----EI---GKRKEUKEDSLWSAKESINNDKLQDDSEMLP 651
      ||| | :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1006 -ANPKKSLLKATAKVGTETTKHLEISTDCKILGVLTKTDTTSMK--LSPDSSSDTSSEN 1062
      ||| | :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY   652 KKLALLTEFRSLVTKNSTSRNPSCINDDYCOLKNFKKKVTYPGAGKLPHIIGGSDLTAH 711
      ::||:::||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db   1063 VRILLTDEQS--VKSSVRSHPSVSTRDSSEER-----KKVRFAGDVGTPTKLNPPPLFGN 1115
      ||| | :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY   712 ---HARKNTLEEWLRQEMEVQNQH-----AKEESLADDLFRYNP 748
      ||| | :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1116 GLQWRSESSLHEKI-NGLTTKSHDDVSIAKRTSLFTDI-EYQP 1158
      ||| | :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT      8
S51441
hypothetical protein YLR309c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein L2142.5
C:Species: Saccharomyces cerevisiae
C:Date: 23-Feb-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C:Accession: S51441
R:Pauley, A.
submitted to the EMBL Data Library, November 1994
A:Description: The sequence of S. cerevisiae cosmid L2142.
A:Reference number: S51437
A:Accession: S51441
A:Molecule type: DNA
A:Residues: 1-911 <PAU>
A:Cross-references: EMBL:U17247; NID:g577216; PIDN:AAB67359.1; PID:g577221;
C:Genetics:
A:Gene: SGD:IMH1
A:Cross-references: SGD:S0004300; MIPS:YLR309c
A:Map position: 12R

Query Match          4.5%; Score 174; DB 2; Length 911;
Best Local Similarity 21.2%; Pred. No. 0.097;
Matches 136; Conservative 113; Mismatches 275; Indels 126; Gaps

```

R3

Db 582 NGKLSERLNIQE---KYNTLQNVKNSNEHIDSIRKQCEBLNVKLESTYKKILSLSEDEL 638

QY 657 TEFRLSVIKNSTGRN-----PSGINDYQGLKNFKFKKVTY--PGAGKLPHIIGSD 708

Db 639 NEYANIYQDQTREANTLRLRVSDSQTDDSSKQKELE--NKLAVLTDEKNKLE-----AEL 691

QY 709 IAHHARKNTELEWLRQEMEVQNO-HA---KEESL-----ADDLFRYNPYLKR 753

Db 692 DLQTSRKATELQEMKHTVTTELKSEIHALKUREGLKSEVDALKHVNNDIRK 743

RESULT 9

I49464

alpha cardiac myosin heavy chain - mouse

C;Species: Mus musculus (house mouse)

C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Feb-2001

C;Accession: I49463; I49463; I49462; I49461; I49604

R;Quinn-Laquer, B.K.; Kennedy, J.E.; Wei, S.J.; Beisel, K.W.

Genomics 13, 176-188, 1992

A;Title: Characterization of the allelic differences in the mouse cardiac alpha-myosin H

A;Reference number: A38207; MUID:92250040

A;Accession: I49464

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1938 <RES>

A;Cross-references: GB:M76601; NID:g191623; PIDN:AAA37162.1; PID:g191624

A;Accession: I49463

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-193, 'D', 195-837, 'S', 839-955, 'N', 957-1938 <RE2>

A;Cross-references: GB:M76600; NID:g191621; PIDN:AAA37161.1; PID:g191622

A;Accession: I49462

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1938 <RE3>

A;Cross-references: GB:M76599; NID:g191619; PIDN:AAA37160.1; PID:g191620

A;Accession: I49461

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-544, 'A', 546-1938 <RE4>

A;Cross-references: GB:M76598; NID:g191617; PIDN:AAA37159.1; PID:g191618

R;Gulick, J.; Subramaniam, A.; Neumann, J.; Robbins, J.

J. Biol. Chem. 266, 9180-9185, 1991

A;Title: Isolation and characterization of the mouse cardiac myosin heavy chain genes.

A;Reference number: I49604; MUID:91225025

A;Accession: I49604

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-67 <RE5>

A;Cross-references: GB:M62404; NID:g192609; PIDN:AAA37424.1; PID:g192610

C;Superfamily: myosin heavy chain; myosin motor domain homology

C;Keywords: ATP; cardiac muscle; heart; nucleotide binding; P-loop

F;88-768/Domain: myosin motor domain homology <MMOT>

F;178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 4.4%; Score 173; DB 2; Length 1938;

Best Local Similarity 22.7%; Pred. No. 0.3;

Matches 92; Conservative 61; Mismatches 122; Indels 130; Gaps 17;

QY 365 ADTESEQADTWLSERPK-----EIKVSKMEQKFRMLSDQAPTVKESCTSSNNNSWS 418

Db 845 AETEEKMANMKBEFRGVKDALEKSEARKKEKMWSLQEKNDLQVQAEQDN----- 899

QY 419 NTLAKMRIPNYQLSPTKLPSTKNSKDRASQQQNTNSIRNYFQPSKKR-----ERDEENQEM 475

Db 900 -----LNDABERCQD-----LKNKIQLEAKVKEMTERLEDEEM 934

QY 476 S----SCKSARIETSCS--LLSQTPATPSLWK--NKEQHSLENEPVDNTSDNNLFTDIDLK 530

Db 935 NAEITAKKRKLEDECSLKKDIDDLTLTAKVEKEKHATENK-----VKN 979

QY 531 IVNSASKSHAAEKLRSNKK-----REMDDVAIEDVEQLFKDTPKPELEIDVKYQK 583

Db 980 LTEMAGLDBIIAKLTKEKKALQEAHQOALDDLOAEEDKVNTL---TKSKVKLEQQVDDL 1036

QY 584 EEDVNVKRRPRMDTETNDTFESDEAVPESSKISQSENEIGCKKRELKEDSLSAKETSN--ND 641

Db 1037 EGSLEQEKKYRMDLER-----AKRKLGLDKLTQESIMDLND 1074

QY 642 KLODDSEMLPKKLLLTETFRSLVINKNSTRNPSGINDYQ-OLKNFKFKPKKVTYPGAGKLP 700

Db 1075 KIQ-----LEEKLKKKEF-----DISQNSKITEDEQALALQKKLKE----- 1112

QY 701 HIIGGSLIAHHARKNTELEEWLRQEMEVQNOHAKESLADDLFR 745

Db 1113 -----NQARIEEELEEAE---RTARAKVEKLRSDLSR 1143

RESULT 10

JC5497

claustrin - chicken

N;Alternate names: keratan sulfate proteoglycan

C;Species: Gallus gallus (chicken)

C;Date: 07-Jul-1997 #sequence_revision 12-Sep-1997 #text_change 21-Jul-2000

C;Accession: JC5497; PC4334; S37561

R;Burg, M.A.; Cole, G.J.

J. Neurobiol. 25, 1-22, 1994

A;Title: Claustrin, an antiadhesive neural keratan sulfate proteoglycan, is structura

A;Reference number: JC5497; MUID:94157526

A;Accession: JC5497

A;Molecule type: mRNA

A;Residues: 1-1038 <BUR1>

A;Cross-references: EMBL:X67778; NID:g406318; PIDN:CAA47988.1; PID:g406319

A;Accession: PC4334

A;Molecule type: protein

A;Residues: 79-83;299-412;485-502 <BUR2>

A;Experimental source: brain

C;Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the ne

C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate

F;267-270/Region: cell attachment (R-G-D) motif

F;112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted

F;152,249,440,793,820/Binding site: chondroitin sulfate (ser) (covalent) #status pred

Query Match 4.4%; Score 172.5; DB 2; Length 1038;

Best Local Similarity 18.6%; Pred. No. 0.14; 326; Indels 309; Gaps 42;

Matches 176; Conservative 135; Mismatches 326; Indels 309; Gaps 42;

QY 6 PAAGPAGGEPYRLTGVVEY---VVRKNKCALLIENDQ-----SISRNHAVLTANFSVTN 56

Db 19 PAA-PPSPSLSHRFLDSKFYLLVWIGE-----LVTEEHLRAIAIERGIRSDMTNLTEN 72

QY 57 LSQTFDEIPVITLKDNSKYGTFTVNEEKWONGFSRTLK-----SGDGIITFGVFGSKFRI- 108

Db 73 LDQ--ELKLFVSRHSARFSPVRGQKILHHRSDVLETVVLINPDSDEAVS-----TEVRLM 125

QY 109 -----EYEPLVACSSCLDVSKTALNQAILQLGGFTVNNWTEECTHLVMVSVKVTI---- 159

Db 126 ITDAARHKLLVLTGQCFENTG-----ELILQSGSFQNFIEFTDQIEGELLSTTHPAN 180

QY 160 KTICALIC-----GRPIVKPEY-----FTEFLKAVQSKKQP 190

Db 181 KASITLFCPEGDKWNSNLDHRNLQDFINIKLNSSSILPEMEGLSEFTEYLS--ESVEVP 238

QY 191 PQIESFYPP-----LDEP-----STGSKNVDLSGROERKQIKFGKFTFI 228

Db 239 SPFDILEPPTSGGFLKSKPCCYIFPGGRGDSRLFAVNGFNMLINGGSEKSCF----- 292

QY 229 FLNAKHKKLSSAVVFGGE-----ARLITEENEEH-----NEFLAPGT 268

Db 293 WKILRHIDRVDSILLTHIGDDNLPGINMSLQRLKTALEEEQSQSGSTNSDMKNLISPD 352

QY 269 CVVDGTG-TNSQTLIPDCQKKWIOSI-----MDMLORQGLRPIPEAEIGLAVIFMTTKN 321

Db 353 GVVFLNYPENLKNMPSFRVK--RSVEENACFTLQYLNKLSMKPEP-----LFRNVGN 402

| | | | |
|----|-----|---|-----|
| QY | 322 | YCDP-----QGHPSGTGLTTPPGPSLSQ-----CVSVD---EKLMPSA----- | 356 |
| | | | |
| Db | 403 | TIDPIILFORMGVGLMEYVNLNPKVSKSKESKYFMHWSGTSKDKAEFLLPNGOEIDPLS | 462 |
| | | | |
| QY | 357 | -----PVTNVTYVADTESQADTWLDSRP----- | 381 |
| | | | |
| Db | 463 | YFTSVSSLIWHHPANPAEKIIRVLPCNSTQYNILEGLEKHLHDLFKOPWTKQDLTCN | 522 |
| | | | |
| QY | 382 | ---KEIKVSKMEQKFRMLSDQAPTVKESCKTSSNNNSMVSNTLAKRIPNYQLSPTKLPS | 438 |
| | | | |
| Db | 523 | IASPAVKQAKLKQRTDSKESLPAKATTTKQDCQKRNLLKHHWSLQSLVQOLEKPKQLES | 582 |
| | | | |
| QY | 439 | INKS---KDRASOOQOTNSIRNYFQSTYK---PERDEENQEMSSCKSARIETSCLLBQT | 493 |
| | | | |
| Db | 583 | KEKTPVKKEKA-----VKPETKIVAEEKDVTTKBEQLGKS---ETSEKQASEK | 627 |
| | | | |
| QY | 494 | QPATPSSLWNKEQHLSENPVDNTSDNNLFTDIDLKSIYKNSAKSHAAEKLRSNKKREM | 553 |
| | | | |
| Db | 628 | QDVKPKVTEK---SVKKEVRAKPEK---KDEKPKVEKSKKEEPLIKKEEKPKK | 679 |
| | | | |
| QY | 554 | DDVAIEDVLEQLFKDTKPELEIDVKKVQOEEDVNVYKRPMDIETNDTFSDEAVPESSK | 613 |
| | | | |
| Db | 680 | ED--IKKEVKKEVKKKEKAEKV-----KEAPPKEAK | 712 |
| | | | |
| QY | 614 | ISOENETGKKRELKEDSLWSAKSEISNNDKLQDDSEMLPKHLILLTFERSLIVKNS'SRN--- | 671 |
| | | | |
| Db | 713 | KEVYKE--EKREIKKEE---KEV-----KKDIKVPKEIKTPTPSTEAKKPAKPKP | 760 |
| | | | |
| QY | 672 | -----PSGINDDYGLKNLKKFKKVTYPGAGKLPHIG----- | 704 |
| | | | |
| Db | 761 | QKKEEPAKKEAVPAGRAKEKGKIKTVKESKYS---GAQTALAANGAAATTVAAVAAAEIT | 818 |
| | | | |
| QY | 705 | --GSDLIAHHARKNT-----ELEEWLROEMEVQNOHAKESLADD | 742 |
| | | | |
| Db | 819 | ASGKELEAERSLSMSSPDLTKDPEELKABEVETIKATRPOVALIED | 864 |

```
QY 705 --CSDLIAHARKNT-----ELEWLROEMEQNHAKESLADD 742
Db 819 ASCKLEAERSLMSPEDLTQDFELKAEVETIKATRPQVALIED 864
```

RESULT 11

A02986 myosin alpha heavy chain, cardiac muscle - rabbit (fragment)
N:Alternate names: alpha isomyosin
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 20-Jun-2000
C:Accession: A02986; I46865
R:Kavinsky, C.J.; Umeda, P.K.; Levin, J.E.; Sinha, A.M.; Nigro, J.M.; Jakovcic, S.; Rabinowitz, M.D. J. Biol. Chem. 259, 2775-2781, 1984
A:Title: Analysis of cloned mRNA sequences encoding subfragment 2 and part of subfragment 1 from rabbit skeletal muscle.
A:Reference number: A02986; MUID:84135762
A:Accession: A02986
A:Molecule type: mRNA
A:Residues: 1-465 <RNA>
A:Cross-references: GB:K02443; NID:g165532; PIDN:AAA31412.1; PID:g165533
A>Note: The authors translated the codon AAG for residue 149 as Arg
R:Sinha, A.M.; Umeda, P.K.; Kavinsky, C.J.; Rajamanickam, C.; Hsu, H. Proc. Natl. Acad. Sci. U.S.A. 79, 5847-5851, 1982
A:Title: Molecular cloning of mRNA sequences for cardiac alpha- and beta-form myosin head domains from rabbit skeletal muscle.
A:Reference number: I46865; MUID:83299886
A:Accession: I46865
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 233-270, 'L', 272-353 <SIN>
A:Cross-references: GB:J00671; NID:g165528; PIDN:AAA31410.1; PID:g165529
C:Comment: In the rabbit the expression of alpha and beta isomyosins follows a defined ontogenetic pattern. The expression of alpha isomyosin is induced by thyroid hormone as well as by other stimuli.
C:Comment: Protease-sensitive functional regions located at the head-rod (S1-S2 subfragment) and actin filament during contraction.
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; cardiac muscle; coiled coil; heart; muscle; thick filament
F:1-36/Domain: head (fragment) <CAR>
F:37-465/Domain: rod (fragment) <ROD>

| | | | |
|----|-----|--|-----|
| QY | 322 | YCDP-----QGHSTGLKTTTPGSPLSQ-----GVSVD---EKLMPESA----- | 356 |
| Db | 403 | TIDPIILFORMGVGLKLEMYLVINPVKSSKESYFMHWSGTSKDKAEFLLPNGOELDPLS | 462 |
| QY | 357 | -----PWNTTYVADTSEQADTWDLSERP----- | 381 |
| Db | 463 | YFTSVSSLIWHHPANPAEKIIRVLFPGNSTQYNILEGLEKLKHLDFLKQPMVTKQDUTGN | 522 |
| QY | 382 | ---KEIKVSKMEQKFRMLSODAPTVKBSCKTSSNNNSMVSNTLAKMRIPNYQLSPTKLPS | 438 |
| Db | 523 | IASPAVQAOKLKORTDSKESILKPAAKTTTKQDCQKRNKKKHWSLQSLVQOLEKPKQLES | 582 |
| QY | 439 | INKS---KDRASOOQOQNTSRNFQPTKK--RRDEENQEMSSCKSARIETSCSLEQOT | 493 |
| Db | 583 | KEKTPVAKKEXA-----VKPETHVAEKDVTYTKEEQLGKS---ETSEKQASEK | 627 |
| QY | 494 | QPATPSLWNKKEQHLSENPVDTNSDNNLTDDTLKSIVKNSASKSHAAEKLRNSKKREM | 553 |
| Db | 628 | QDVKPKYTKK-----SVKVKAKPEK---KDEKPKPEKSKKEEKLPIIKKEPKPK | 679 |
| QY | 554 | DDVAIEDVLEQLFKDTKPELEIDVKYQKQPEDVNVKRPRMDIETNDTFSDEAVPSSK | 613 |
| Db | 680 | ED--IKKEVKKEVKKEKKEAKKEV-----KEAPPKEAK | 712 |
| QY | 614 | ISOENETCKRELKEDSLWAKELSNNDKLQDDSEMLPKLLLTFFRSILVKNSTSRN-- | 671 |
| Db | 713 | KEVKE--EKRETKKEE---KEV-----KKDIKVPKEIKKTPPTPSTBAKPAAPKPK | 760 |
| QY | 672 | -----PSGINDDYGLKNFKKFKKVPYPGAKLPHIIG----- | 704 |
| Db | 761 | OKKEEPAKKEAVPAGKAKKEGKIKTVKESKVS--GAQTALAAGAAATTVAAVAAAEIT | 818 |
| QY | 705 | --GSDLIAHARKNT-----ELEEWLRQEMEVQONHAKESLADD | 742 |
| Db | 819 | ASGKELEAERSLSMPDLTKDFEELKAAEVEVETIKATRPQVALIED | 864 |

```
QY 705 --CSDLIAHARKNT-----ELEWLROEMEVQNHAKESLADD 742
Db 819 ASCKLEAERSLMSPEDLTQDFELKAEVETIKATRPQVALIED 864
```

Db 902 DAERCQ-----LIKKIOLEAKVKRETERLEDEEEMNAELTS-KRKLEDECESEUKD 955
QY 493 TOPATPSLWK-NKEQHLSENPEVDTSNNLFTDILKSIVKNSASKASHAAELRSNKK- 550
Db 956 IDLELTAKVEREKHATENK-----VKNLTEEMAGLDEIIAKLTREKKA 1000
QY 551 -----REMDVAIEDVLQOLFKDTPKELEIDVVKQKQBEDVNVRRPRMDIETNDTFS 604
Db 1001 LOEAHQALDDLOAEEDKVNTL---TKSKVKLEQQVDLEGSLEQEKVKRMDLER----- 1052
QY 605 DEAVPESKISQENIGKKEKRELKEDSLWSAKEISN---NDKLQDDSEMLPKLLITPRSL 662
Db 1053 -----AKRKLEGLDUNVTOESIMDLENDKIQ-----LEELKKKEF--- 1087
QY 663 VIKNSTSRNPSSGINDDYG-OLKNFKFKPVYTPGAGKLPHIIGSGDLIAHARKNTELEE 721
Db 1088 ----DISQONSKIEDQALALQKKLKE-----NQARIELEEE 1122
QY 722 WLROEMEVQNHAKESLADDLFR 745
Db 1123 ELEAE---RTARAKVEKLRSDLTR 1143

RESULT 13

ORMSPI

microtubule-associated protein MAP1B - mouse
N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protein
C:Species: Mus musculus (house mouse)
C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 01-Sep-2000
C:Accession: S07549; S44387; A33645
R:Noble, M.; Lewis, S.A.; Cowan, N.J.
J. Cell Biol. 109, 3367-3376, 1989
A:Title: The microtubule binding domain of microtubule-associated protein MAP1B contains
A:Reference number: A33645; MUID:90094539
A:Accession: S07549
A:Molecule type: mRNA
A:Residues: 1-2464 <NO>
A:Cross-references: EMBL:X51396; NID:g52999; PIDN:CAA35761.1; PID:g53000
R:Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.
Arch. Biochem. Biophys. 310, 428-432, 1994
A:Title: Binding of heat-shock protein 70 (hsp70) to tubulin.
A:Reference number: S44387; MUID:94234720
A:Accession: S44387
A>Status: preliminary
A:Molecule type: protein
A:Residues: 653-663, 'IC' <SAN>
C:Superfamily: microtubule-associated protein MAP1B
C:Keywords: microtubule binding; phosphoprotein; tandem repeat
F:589-786/Domain: microtubule binding #status experimental <MTB>
F:589-592,639-642,649-652,655-658,660-663,668-671,674-677,679-682,683-686,687-690,691-69
R-K-E/D-X)
F:1861-2064/Region: 17-residue repeats
F:91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/binding site: ph
F:147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (co
F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 4.4%; Score 170; DB 1; Length 2464;
Best Local Similarity 18.0%; Pred. No. 0.6;
Matches 171; Conservative 148; Mismatches 301; Indels 330; Gaps 44;

QY 6 PAAGPAGGEPYRLLTGVYVGRKNCAILIENDQISNRNH---AVLTANFSV-----TNLS 58
Db 20 PAASTPSLSHRELDISKFY-----LLVVGVETVEEHLRRAIGNIELGIRSDWNL 71
QY 59 QTD---EIPVLITKDNKSYGTFVNEEKMQNGFSRTLK-----SGDITFGVFGSKFRI 108
Db 72 ECNLDQELKLFVSRHARSPEYPGQKILHRSDVLETVVVLINPSDEAVS-----TEVRL 126
QY 109 -----EYELPVACSSCLDVSQKLTALNOAILQLGGFTVNNWTECHLVNVSVKVI--- 159
Db 127 MITDAARHLLVLTGCFENFG-----ELILQSGSFSFQNFIFITDQEI GELLSTHPA 181

QY 160 -KTICALIC-----GRPIVKPEY-----FTEFLKAVQSKKQ 189
Db 182 NKASLTLCFCEEGDWKNSNDRHNLQDFINIKLNASILPEMEGLSEFTEYLS--ESVEV 239
QY 190 PPOIESYPP-----LDEP-----SIGSKNVDLSGRQERKQIFKGTFF 227
Db 240 PSPFDILEPTSGGFLKLSKPCCYIFPGRGDSALFAVNGFNMLINGSERKSCF----- 294
QY 228 IFLNAKHKKLSAVVFGGE-----ARLITEENEENFFFLAPGTCVVVDGTGIN 277
Db 295 -WKLIRHLDRVDSILLTHIGDDNLPGINSMLQKIAELEER-----SQSSTS 341
QY 278 SQTLLIPDCQKKWISIMDLQROGLRPIPEAIGLAVIFMTT-KNYCDPQGHPTGLKTT 336
Db 342 N-----SDWMKNLIS-----PD-----LGVVFLNVPENLKDPE----- 369
QY 337 TPGPSLSQGSVDKELMPSAPVNTTIVYADTSEQADTWLSERP----- 382
Db 370 ---PNIKMKRSIEEACFTLOYLN-----KLSMKPEPLFRSVGNWTIEPVI 410
QY 383 -----EIKVSKMEQKFRMLSQDAPTVKESCKTSSNNNSMV----- 417
Db 411 LFOKMVGKLEMYVLNVPVSKSEMQYF--MQOWTGTNKKDAELLIPNGQVDDIPISYLS 468
QY 418 -----SNTAKMRIPNYQLSPTTKLPSINKS-----DRASQOQQTNSI--- 455
Db 469 VSSLIVWHPANPAEKIIRVLPFGNSTQYNILEGLEKHLKHLDFLKQPLATOKDLTGQVPTP 528
QY 456 -----RNYPOPSTK-----KREDEENQEMSSKSARLETSCSLEQTQP 495
Db 529 PVQVKLIKQADRESLSUKPATKPVASKVRKSEKEETPEVT--KTSQVE----- 575
QY 496 ATPSLMKNKQHLSENPEVDTSNNLFTDILKSIVKNSASKASHAAELRSNKKREM-D 554
Db 576 KTPKVESKEKVLVKKDKPVKTESKPSV-TEKEYSSKEEQSPVKAEEVAEKQATESKPVTK 634
QY 555 DVAIEDSV---LEQ-----LFDKTPK---ELEIDVKVQ-KQEDVNV 589
Db 635 DKVVKKEIKTKLEKKEKPKKEVVKKEDKTPLLKDKPKRKEEVKKEIKKEIKERREL 694
QY 590 RKPRMDIETNDTFSDEAVPESKISQENIGKKEKRELKEDSLWSAKESINNDKLOD---D 646
Db 695 KKEVKETPLKDA-KKEVKKEKKEKKEKPEKKEIKKIS-----KDKIKSTPOSSTKPK 749
QY 647 SEMLPKLLLTFRSLVIKNSTSRNP--SGINDDYGOLKNFKFKKVTYPG-----AGK 698
Db 750 SALKPKVAKKEE-----STKKEPLAAGKLKDKGKVKVKKKEGKTEAATAVGTAA 801
QY 699 LPHIIGSDILIAHARKNTELEEWLROEME-----VONQHAKESLADDL 743
Db 802 TAAVVAAGAAAGSPVKVELEASRLMSSPEDLTAKDFEELKAEEDVAKDI 851

RESULT 14

S20117

protein kinase BCK1 (EC 2.7.1.1) - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein J0906; protein kinase SLK1; protein kinase SSP31; protein
C:Species: Saccharomyces cerevisiae
C>Date: 23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change 24-Sep-1999
C:Accession: S20117; S50298; S22285; S19061; JQ1432; S56872; S30794; JQ1118
R:Costigan, C.; Gehring, S.; Snyder, M.
Mol. Cell. Biol. 12, 1162-1178, 1992
A:Title: A synthetic lethal screen identifies SLK1, a novel protein kinase homolog im
A:Reference number: S20117; MUID:92186847
A:Accession: S20117
A:Molecule type: DNA
A:Residues: 1-1478 <COS>
A:Cross-references: EMBL:M84389
A:Experimental source: strain S288C
R:Mosga, T.; Boles, E.; Schaeff-Gerstenschlaeger, I.; Schmitt, S.; Zimmermann, F.K.
Yeast 10, 1481-1488, 1994
A:Title: Sequence and function analysis of a 9.74 kb fragment of Saccharomyces cerevi


```
Db 1321 AVVSNGLGQYHH-----MSMSLTEAE--VFSCNR-----1347
QY 301 GLRPIPAEIGLAVIFMTTKNYCD-----POGHPSTGLKTTTP--GPSLSQGVSVDEK 351
Db 1348 -----FOGRISYCSDMWEYKPOS--LEALNPTFPKIGTCTPSSVTVSMADN 1392
QY 352 LMPSPAVNTTYYVADTESQADTWLSERPKEIKVSRMEQFRMLSDQAPTAVKESCKTSS 411
Db 1393 LSGILELHSSFDLTNTTSKSVNPTPFGEYPTMI-----WEAPKL---CTSDN 1436
QY 412 NNSWSVN-----TLAK--MRIPNYQLSPKLPSPINKSKDRASQQOQTN---453
Db 1437 NNNNSCKNDQORNEEKVDKTRSLIKLSKIANVYFGPFPSAITKNSNFAIKPLEINEQP 1496
QY 454 SIRNYFOPSTK-----KREDEENQEM-----SSCKSARIETSCSLLEOTQOPATPSL 500
Db 1497 DIINCNPVSKSAVSKTNLKNVSNKYKQKMHFAKSHFHPADLEMPADLEISLMRT---1553
QY 501 WKNKEQHLSNEPVDT--NSDNNLFTDTDLKSIVKNSASKSHAAEKLRSNKKREMDDDVAI 558
Db 1554 -NNTESNLKNTKTIATISINEMMDMISDDNVIPMLQNNVAK-----HQKRKVEICSL 1603
QY 559 EDEVLEQLFKDTPELEIDVKVQOEEDVNVKRPMDIETNDFSD-----605
Db 1604 DDNVVYTSRKLTKRELLKCS-KTEKPDYEPSSAKTPQGNFFKGDSSAKIDLFGSHKFIKD 1662
QY 606 ---EAVPESSKISOENEIGKKRELKEDSL-----WSAKEI-----637
Db 1663 ERAETCPDKTEIPEEYIGSTQNLHENKAIKDKVVPINANNLKQKNVKEIPTTVDVGQLP 1722
QY 638 -----SNNDKLODDSEMLPKLL-LTEFR-----SLVIKNSTSRNPSPGINDDYG-----680
Db 1723 ILDYNYSNSTRNENLFSKVFGEKNEYQYGFPMGKLIERMGTGTGPKVYVSEYSSSPIA 1782
QY 681 -----OLKNFKFKKVTYP-----GAGKLPHIIGGSDLIAHHARKNTELEEWL 723
Db 1783 AIFSPKADSHEFADSNKLOKTTKPSICMKFGNCGRLPDL-----EPNKN-----WL 1829
QY 724 ROEM-EVQNOHA-----KEESLADDLF 744
Db 1830 RONNAQFAGHHASVIDSKSFTKLKETSIVFQWF 1863
```

Search completed: August 15, 2002, 01:42:27
Job time: 5262 sec

| Result No. | Score | % | | Length | DB | ID | Description |
|------------|-------|-------|-------|--------|------------|--------|--------------|
| | | Match | Query | | | | |
| 1 | 179.5 | 4.6 | 1790 | 1 | US01_YEAST | P25386 | saccharomyc |
| 2 | 173 | 4.4 | 1938 | 1 | MYH6_MOUSE | P02566 | mus musculus |
| 3 | 171 | 4.4 | 465 | 1 | MYH6_RABIT | P04460 | oryctolagus |
| 4 | 171 | 4.4 | 1939 | 1 | MYH6_MESAU | P13539 | mesocricetu |
| 5 | 170.5 | 4.4 | 1939 | 1 | MYH6_HUMAN | P13533 | homo sapien |
| 6 | 170.5 | 4.4 | 2492 | 1 | ATRX_HUMAN | P46100 | homo sapien |
| 7 | 170 | 4.4 | 2464 | 1 | MAPB_MOUSE | P14873 | mus musculus |
| 8 | 169.5 | 4.3 | 1478 | 1 | BCKL_YEAST | Q01389 | saccharomyc |
| 9 | 168.5 | 4.3 | 1233 | 1 | YF16_YEAST | P43597 | saccharomyc |
| 10 | 168.5 | 4.3 | 3672 | 1 | LM12_CAEEL | Q21313 | caenorhabdi |
| 11 | 167.5 | 4.3 | 2349 | 1 | TPR_HUMAN | P12270 | homo sapien |
| 12 | 166 | 4.3 | 1359 | 1 | ATRX_CAEEL | Q9U7e0 | caenorhabdi |
| 13 | 166 | 4.3 | 1935 | 1 | MYH7_HUMAN | P12883 | homo sapien |
| 14 | 165 | 4.2 | 667 | 1 | CYL1_BOVIN | P35662 | bos taurus |
| 15 | 165 | 4.2 | 1938 | 1 | MYH6_RAT | P02563 | rattus norv |
| 16 | 165 | 4.2 | 1939 | 1 | MYH1_HUMAN | P12882 | homo sapien |
| 17 | 164.5 | 4.2 | 2871 | 1 | DESP_HUMAN | P15924 | homo sapien |
| 18 | 164 | 4.2 | 915 | 1 | ZDSL_YEAST | P50111 | saccharomyc |
| 19 | 164 | 4.2 | 1381 | 1 | YB7_YEAST | P34216 | saccharomyc |
| 20 | 163 | 4.2 | 952 | 1 | YK15_CAEEL | P46012 | caenorhabdi |
| 21 | 162.5 | 4.2 | 539 | 1 | YAH2_YEAST | P39705 | saccharomyc |
| 22 | 162.5 | 4.2 | 1427 | 1 | REST_HUMAN | P30622 | homo sapien |
| 23 | 162.5 | 4.2 | 1957 | 1 | YD86_SCHPO | Q10411 | schizosacch |
| 24 | 162 | 4.2 | 1312 | 1 | RA50_YEAST | P12753 | saccharomyc |
| 25 | 162 | 4.2 | 2748 | 1 | NUM1_YEAST | Q00402 | saccharomyc |
| 26 | 161 | 4.1 | 1928 | 1 | MYSL_YEAST | P08964 | saccharomyc |
| 27 | 160 | 4.1 | 1934 | 1 | MYH7_MESAU | P13540 | mesocricetu |
| 28 | 159.5 | 4.1 | 1189 | 1 | YH76_YEAST | P47035 | saccharomyc |
| 29 | 158.5 | 4.1 | 1664 | 1 | INT1_CANAL | P53705 | candida alb |
| 30 | 158 | 4.1 | 1744 | 1 | TANA_XENLA | Q01550 | xenopus lae |
| 31 | 158 | 4.1 | 1938 | 1 | MYH4_RABIT | Q28641 | oryctolagus |
| 32 | 157.5 | 4.0 | 1937 | 1 | MYH8_HUMAN | P13535 | homo sapien |
| 33 | 157 | 4.0 | 1164 | 1 | BAG_STRAG | P27951 | streptococc |

```
KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 4.6%; Score 179.5; DB 1; Length 1790;
Best Local Similarity 23.1%; Pred. No. 0.065;
Matches 100; Conservative 77; Mismatches 168; Indels 87; Gaps 19;

QY 378 SERPKETKYSKQKFMKLSODAPTVKESCKTSSNNNSMVSNTLAKMRIPYQLSPTKLP 437
Db 988 SKNESSIQLSNLQNKIDMSQEKENFQIERGSIKIEQLKKTIS-----DLEQTKEE 1040

QY 438 SINK-----SKDRASQ-----OQTNSIRNVFOPSTKKRDERDEENQMSCKSARIE 484
Db 1041 IISKSDSSKDEYESQISLLKEKLETATTANDENVNKISELTKTREELEAELAYKNLKN 1100

QY 485 TSCSLLEQTOPATPSLWKNKEQHL-----ENEPVDNTSDNNLFTDTDLKSIK 535
Db 1101 LE-TKLETSEKALKEV-KENEELKKEEIQLEKEATETKQOLNSLR-ANLESEKEHEDL 1157

QY 536 ASKSHAAEKLRSNKKREMD--VAIEDEVL-----EQLFKDTKPELEIDVKKVK--QEED 586
Db 1158 AALQKYEQIAKQYNEIEISQNLDEITSTQOENESIKKNDLEGEVKAAMKTSSEQ 1217

QY 587 VNVKRP-----RMDIETNDTFSDEAVPESKISQENEIGKKRELKEDSLWSAK 635
Db 1218 SNLKSEIDALNLQIKELKKNETNEA-----SLESIK-SVESETVKIKELQDECNFKK 1272

QY 636 EISN-NDKLQDDSEMLPKLLLTFFSLVTKNSTSRNPISINDYQOLKNFKFKKVTYP 694
Db 1273 EVSELEDKLR-ASEDKNSKYLEQKSEKIEKELDAKTTTELKTOLEKITNLKSAKES 1331

QY 695 GAGKLPHIIGGSDLIHARKNTELE-EWLRQEMEQVQNHAKEE----- 737
Db 1332 ELSRLKK-----TSSEERKNAEEQLEKLNQIKNQAFERKLLNKGSGSTITQEYS 1384

QY 738 -----SLADDLFR 745
Db 1385 EKINTLEDELIR 1396

RESULT 2
MYH6_MOUSE STANDARD; PRT; 1938 AA.
AC Q02566; Q64258; Q64738;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle alpha isoform (MYH6-alpha).
GN MYH6 OR MYHCA.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C, A/J, C57BL/6J, AND DBA/2J;
RC MEDLINE=92250040; PubMed=1577481;
RA Quinn-Laquer B.K., Kennedy J.E., Wei S.J., Beisel K.W.;
```

```
RT "Characterization of the allelic differences in the mouse cardiac
RT alpha-myosin heavy chain coding sequence.";
RL Genomics 13:176-188(1992).
RN [2]
RP SEQUENCE OF 1-67 FROM N.A.
RC STRAIN=AKR;
RX MEDLINE=91225025; PubMed=2026617;
RA Gulick J., Subramanian A., Neumann J., Robbins J.;
RT "Isolation and characterization of the mouse cardiac myosin heavy
RT chain genes.";
RL J. Biol. Chem. 266:9180-9185(1991).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE. SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER RE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M76598; AAA37159.1; -
DR EMBL; M76599; AAA37160.1; -
DR EMBL; M76600; AAA37161.1; -
DR EMBL; M76601; AAA37162.1; -
DR EMBL; M62404; AAA37424.1; -
DR HSSP; P08799; LMMD.
DR SWISS-2DPAGE; Q02566; MOUSE.
DR MGD; MGI:97255; Mvha.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family;
KW Calmodulin-binding; Polymorphism.
FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
FT DOMAIN 783 812 IQ.
FT DOMAIN 842 1938 COILED COIL (POTENTIAL).
FT NP_BIND 178 185 ATP.
FT DOMAIN 657 679 ACTIN-BINDING.
FT DOMAIN 759 807 ACTIN-BINDING.
FT DOMAIN 816 833 CALMODULIN-BINDING (BY SIMILARITY).
FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 697 697 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 707 707 ALKYLATION (SH-2) (POTENTIAL).
```

| | | | | |
|---|----------|---|--------------------------------------|-------------------------|
| FT | VARIANT | 194 | 194 | Y -> D. |
| FT | VARIANT | 545 | 545 | S -> A. |
| FT | VARIANT | 838 | 838 | I -> S. |
| SQ | SEQUENCE | 1938 AA; | 223564 MW; | EAD789ADAG6818FB CRC64; |
| Query Match 4.4%; Score 173; DB 1; Length 1938; | | | | |
| Best Local Similarity 22.7%; Pred. No. 0.16; | | | | |
| Matches 92; Conservative 61; Mismatches 122; Indels 130; Gaps | | | | |
| Qy | 365 | ADTSEQADTWDLSERPK---- | EIKYSKW6KFRMLSQDAPTVKESCCKTSSNNNSMVS | 418 |
| Db | 845 | AETEEMANNKEEFGVKDALEKSEARKEEMVSLQ6KNLDLQVQAEQN----- | 899 | |
| Qy | 419 | NTLAKMRIPNYQLSPTKLPISNKSDRASQOQOTNSIRNYFQPSPTKKR-- | ERDEENQEM | 475 |
| Db | 900 | -----LNDAEERCDQ----- | LINKKIQLEAKYKEMTERLEDEEEM | 934 |
| Qy | 476 | S----SCKSARIETSCS-LLEQQTAPTPSLWK-NKEQHLSENEPVDNTSDNNLFTDTDLKS | 530 | |
| Db | 935 | NAELTAKRKRLDECSELKDKDIDLELTAKVEKEKHATENK----- | VKN | 979 |
| Qy | 531 | IVKNSASKSHAAEKLRSNKK----- | REMDVALEDEVLEQLPKDTKPELEIDVKYQKQ | 583 |
| Db | 980 | LTEEMAGLDEITIAKLTKEKKALQEAHQHQAALDQLAAEDKQVNTL--- | TKSKVKLEQQVDDL | 1036 |
| Qy | 584 | EEDVNVKRPRMDIETNDTFSDEAVPESSKISQENEICKRELKEDSLWSAKEISN-ND | 641 | |
| Db | 1037 | EGSLEQEKVYMDLER----- | AKRLEGDLKLTOESTWDLND | 1074 |
| Qy | 642 | KLQDDSEMLPKLLLTFFRSLVIKNSTSRNPSPGINDDYG-QLKNFKPKPKVYTPGAKLKP | 700 | |
| Db | 1075 | KLQ-----LEEKLKKEF----- | DISQONSKIEDQALALQJLKLKE----- | 1112 |
| Qy | 701 | HIIGSGDLTAHHARKNTELEEWLRMEVQNOHAKESLADDLFR | 745 | |
| Db | 1113 | -----NOARTELEEELEAE----- | RTARAKVEKLRSDLR | 1143 |

| | | | | | |
|------------|---|-----------------------------------|------|---------|--|
| RESULT | 3 | | | | |
| MYHC_RABIT | | | | | |
| ID | MYHC_RABIT | STANDARD; | PRT; | 465 AA. | |
| AC | P04460; | | | | |
| DT | 13-AUG-1987 | (Rel. 05, Created) | | | |
| DT | 13-AUG-1987 | (Rel. 05, Last sequence update) | | | |
| DT | 16-OCT-2001 | (Rel. 40, Last annotation update) | | | |
| DE | Myosin heavy chain, cardiac muscle alpha isoform (MyHC-alpha) (Alpha | | | | |
| DE | isomyosin) (Fragment). | | | | |
| DE | GN | MYH6. | | | |
| OS | Oryctolagus cuniculus (Rabbit). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus. | | | | |
| OX | NCBI_TaxID=9986; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE=84135762; PubMed=6321481; | | | | |
| RA | Rakovsky C.J., Umeda P.K., Levin J.E., Sinha A.M., Nigro J.M., | | | | |
| RA | Jakovcic S., Rabinowitz M.; | | | | |
| RT | "Analysis of cloned mRNA sequences encoding subfragment 2 and part of | | | | |
| RT | subfragment 1 of alpha- and beta-myosin heavy chains of rabbit | | | | |
| RL | heart."; | | | | |
| RL | J. Biol. Chem. 259:2775-2781(1984). | | | | |
| RL | [2] | | | | |
| RP | SEQUENCE OF 91-177 FROM N.A. | | | | |
| RX | MEDLINE=84221901; PubMed=6328491; | | | | |
| RA | Friedman D.J., Umeda P.K., Sinha A.M., Hsu H.J., Jokovic S., | | | | |
| RA | Rabinowitz M.; | | | | |
| RT | "Characterization of genomic clones specifying rabbit alpha- and | | | | |
| RT | beta-ventricular myosin heavy chains"; | | | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 81:3044-3048(1984). | | | | |
| CC | -1- FUNCTION: MUSCLE CONTRACTION. | | | | |
| CC | -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 | | | | |
| CC | HEAVY CHAIN SUBUNITS (MHC), 2 ALKALIC LIGHT CHAIN SUBUNITS (MLC) | | | | |

```

CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -I- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -I- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPEITIVELY, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -I- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (SI) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcements
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02443; AAA31412.1; -.
CC DR EMBL; K01697; AAA31416.1; -.
CC DR PIR; A02986; A02986.
CC DR HSP; P13538; 2MYS.
CC DR InterPro; IPR002928; Myosin_tail.
CC DR Pfam; PF01576; Myosin_tail; 1.
CC KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC FT ATP-binding; Multigene family.
CC FT NON_TER 1 1
CC FT DOMAIN <1 35 MYOSIN HEAD-LIKE.
CC FT DOMAIN 36 >465 COILED COIL (POTENTIAL).
CC FT NON_TER 465 465
CC SQ SEQUENCE 465 AA; 54375 MW; BAC86AFF8CE8E29C CRC64;
CC -----
Query Match 4.48; Score 171; DB 1; Length 465;
Best Local Similarity 21.38; Pred. No. 0.033;
Matches 81; Conservative 67; Mismatches 136; Indels 96; Gaps 14;
Qy 384 IKVKMEQKFRMLSDAPTVKESCKTSNNNSMVSNTLAKRIPNYQLS---PTKLPISN 440
Db 37 LKSAEAKEMAAMKEEFGRIKESLEKSEARRKELEEKMVSLLOEKNDLQLQVQAEQDNLN 96
Qy 441 KSKDRASQQOQTNSIRNVPQSTKKR---ERDENQEMS---SCKSARIETSCS--LLEQT 493
Db 97 DAERCDQ-----LIKNIQLEAKYKVENNERLEDEEENAEELTAKRKLEDECSELKKDI 151
Qy 494 QPATPSLWK-NKEOHLSENPVDNDSNNLFTDCLKSVKNSASKSHAAEKLRSNKK-- 550
Db 152 DDLEULTAKVKEKHATENK-----VKNLTEMAGLDEIIAKLTKEKKAL 196
Qy 551 -----REMDVAIEDVLEQLFKDTKPELEIDVKYQKEEDVNVKRRPRMDIETDFTSD 605
Db 197 QEAHQALDDLAQEDKVNLT--TKAKLKEQQVDLLEGSLEQKVKRMDLER----- 247
Qy 606 EAVPESSKISQENIGKGRKELKEDSLNSAKSIEISNNDKLODDSEMLPKLLLTPEPSLVIK 665
Db 248 -----ARKLEGDLUKLTQESIMD---LENDKQLEERLKKKEF----- 282
Qy 666 NSTSRNPSPGINDYQGLKNFKFKKVTYPGAGKLPHITGGSLTAAHARKNTELEEWLRQ 725
Db 283 -DISQLNSKIEDEQALVLQLQKKL-----ENQARTEELEEELEA 321
Qy 726 EMEVONQAHAKESLADDLFR 745
Db 322 E---RTARAKVEKLRSDLSR 338
CC -----
RESULT 4
MYH6_MESAU
ID MYH6_MESAU STANDARD; PRT; 1939 AA.
AC P13539; Q60562;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

```



```

Db 1118 EEELEEAE---RTARAKVEKRSLSR 1143
||||| | | | | | | | | |
RESULT 6
ATTRX_HUMAN
ID ATTRX_HUMAN STANDARD; PRT: 2492 AA.
AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE nuclear protein) (XNP) (Znf-HX).
GN ATRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND VARIANT S-1860,
RP AND VARIANTS ATR-X.
RX MEDLINE-97123494; PubMed=8968741;
RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
RA Gibbons R.J.;
RT "ATRX encodes a novel member of the SNF2 family of proteins; mutations
RT point to a common mechanism underlying the ATR-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX MEDLINE-97386582; PubMed=9244431;
RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
RA Collea L., Schwartz C., Fontes M.;
RT "Determination of the genomic structure of the XNP/ATRX gene encoding
RT a potential zinc finger helicase.";
RL Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 860-2492 FROM N.A.
RX MEDLINE-95179111; PubMed=7874112;
RA Stayton C.L., Dabovic B., Gulisano M., Geetz J., Broccoli V.,
RA Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA Bianchi M.E., Consalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RL Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE-94214473; PubMed=8162050;
RA Geetz J., Pollard H., Consalez G., Villard L., Stayton C.L.,
RA Millaesau P., Khrestchatsky M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.";
RL Hum. Mol. Genet. 3:39-44(1994).
RN [5]
RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
RX MEDLINE-95211835; PubMed=7697714;
RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
RL Cell 80:837-845(1995).
RN [6]
RP SEQUENCE OF 1375-2492 FROM N.A.
RA Pearce A., Chapman J.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP EZH2 BINDING.
RX MEDLINE-98167853; PubMed=9499421;
RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,
RA Collea L.;
RT "Specific interaction between the XNP/ATRX gene product and the SET
RT domain of the human EZH2 protein.";
RL Hum. Mol. Genet. 7:679-684(1998).
RN [8]
RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC

```

```

RP HETEROCHROMATIN.
RX MEDLINE-20040663; PubMed=10570185;
RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT "Localization of a putative transcriptional regulator (ATRX) at
RT pericentromeric heterochromatin and the short arms of acrocentric
RT chromosomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
RN [9]
RP DISEASE.
RX MEDLINE-20213147; PubMed=10751095;
RA Villard L., Fontes M., ades L.C., Geetz J.;
RT "Identification of a mutation in the XNP/ATRX gene in a family
RT reported as Smith-Fineman-Myers syndrome.";
RL Am. J. Med. Genet. 91:83-85(2000).
RN [10]
RP VARIANT ATR-X SER-1713.
RX MEDLINE-97196774; PubMed=9043863;
RA Villard L., Lacombe D., Fontes M.;
RT "A point mutation in the XNP gene, associated with an ATR-X phenotype
RT without alpha-thalassemia.";
RL Eur. J. Hum. Genet. 4:316-320(1996).
RN [11]
RP VARIANT JM GLN-2131.
RX MEDLINE-96224392; PubMed=8630485;
RA Villard L., Geetz J., Mattei J.-F., Fontes M., Saugier-Verber P.,
RA Munnich A., Lyonnet S.;
RT "XNP mutation in a large family with Juberg-Marsidi syndrome.";
RL Nat. Genet. 12:359-360(1996).
RN [12]
RP VARIANTS ATR-X.
RX MEDLINE-97467722; PubMed=9326931;
RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asenbauer B.,
RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppeler K., Kurosawa K.,
RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,
RA Higgs D.R.;
RT "Mutations in transcriptional regulator ATRX establish the functional
RT significance of a PHD-like domain.";
RL Nat. Genet. 17:146-148(1997).
RN [13]
RP VARIANT ATR-X LEU-246.
RX MEDLINE-20123052; PubMed=10660327;
RA Fichera M., Romano C., Castiglia L., Failla P., Ruberto C., Amata S.,
RA Greco D., Cardoso C., Fontes M., Ragusa A.;
RT "New mutations in XNP/ATRX gene: a further contribution to
RT genotype/phenotype relationship in ATR/X syndrome.";
RL Hum. Mutat. 12:214-214(1998).
RN [14]
RP VARIANT SHS LYS-1742.
RX MEDLINE-99347960; PubMed=10417298;
RA Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,
RA Prieto F., Fontes M., Martinez F.;
RT "Mutation of the XNP/ATRX gene in a family with severe mental
RT retardation, spastic paraplegia and skewed pattern of X inactivation:
RT demonstration that the mutation is involved in the inactivation
RT bias.";
RL Am. J. Hum. Genet. 65:558-562(1999).
RN [15]
RP VARIANT CWS THR-2050.
RX MEDLINE-99326051; PubMed=10398237;
RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
RA Curtis M.;
RT "Carpenter-Waziri syndrome results from a mutation in XNP.";
RL Am. J. Med. Genet. 85:249-251(1999).
RN [16]
RP VARIANTS ATR-X E-175; 178-V-K-198 DEL; S-190; P-219; L-246 AND C-249.
RX MEDLINE-99219535; PubMed=10204841;
RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Belougne J.,
RA Lossi A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M.,
RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;
RT "Evaluation of a mutation screening strategy for sporadic cases of
RT ATR-X syndrome.";

```


CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
 CC OF MAP1B.
 CC -1- SIMILARITY: TO MAP1A.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X51396; CAA35761.1; -.
 CC PIR: S07549; QRMSPL.
 CC MGD: MGI:1306778; Mtap1b.
 CC InterPro: IPR000102; MAP1B_neuraxin.
 CC Pfam: PF00414; MAP1B_neuraxin; 10.
 CC PROSITE: PS00230; MAP1B_NEURAXIN; 7.
 CC Microtubules; Repeat; Phosphorylation.
 CC CHAIN ? 2464 MAP1 LIGHT CHAIN LCL.
 CC FT REPEAT 1874 1890 MAP1B 1.
 CC FT REPEAT 1891 1907 MAP1B 2.
 CC FT REPEAT 1908 1924 MAP1B 3.
 CC FT REPEAT 1925 1941 MAP1B 4.
 CC FT REPEAT 1942 1958 MAP1B 5.
 CC FT REPEAT 1959 1975 MAP1B 6.
 CC FT REPEAT 1993 2009 MAP1B 7.
 CC FT REPEAT 2010 2026 MAP1B 8.
 CC FT REPEAT 2027 2043 MAP1B 9.
 CC FT REPEAT 2044 2060 MAP1B 10.
 CC FT DOMAIN 589 787
 CC LYS-RICH (HIGHLY BASIC, CONTAINS MANY
 CC KKEE AND KKEI/V REPEATS).
 CC SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFD8DA87 CRC64;

Query Match 4.4%; Score 170; DB 1; Length 2464;
 Best Local Similarity 18.0%; Pred. No. 0.3;
 Matches 171; Conservative 148; Mismatches 301; Indels 330; Gaps 44;
 QY 6 PAAGPAGEPYRLITGVYVGRKNCALIEDQISIRNH---AVLTANFSV-----TNLS 58
 Db 20 PAASTSLSHRFLDSKFY-----LLVVGVETVTEEHLRAIGNIELGIRSWDNL 71
 QY 59 QTD---BIPVLTKDNKSKYGTFFNEEKMQNGFSRTLK-----SGDITGVFGSKPRI 108
 Db 72 ECNLDQELKLFVSRHSRSPVPGQKILHRSVDLETVVVLINPSDEAVS-----TEVRL 126
 QY 109 -----EYEPVLVACSSCLDYSGKTALNQAILQLGGFTVNNWTECTHLVWVSVKVTI--- 159
 Db 127 MITDAARHKLVLVTGQCFENTG-----ELIQSGSFSFQFIETFDQIEIGELLSTTHPA 181
 QY 160 -RTICALIC-----GRPIVKPEY-----FTBFLKAVQSKKQ 189
 Db 182 NKASLTLCFPEGDKWNSLDRHNLQDFINIKLSASILPEMEGLSEFTEYLS--ESVEV 239
 QY 190 PQIESYPP-----LDEP-----SIGKNVDLSGRQKQIFKQKTF 227
 Db 240 PSPFDILEPPTSGGFLKSPCCYIFPGGRGDSALFAVNGFNLINGGSKSCF----- 294
 QY 228 IFLNAQKHKLSSAVFVGGE-----ARLITEENEEHNFFLAPGTCVVDITGN 277
 Db 295 -WKLIRHLDRVDSLTHLTHIGDDNLPGINSMLQRIAELEER-----SQGSTS 341
 QY 278 SQTLPDCQKWIQSIQMDLQRLRPIPEAIGLAVIFMTT--KNYCDPQHPSTGLKTT 336
 Db 342 N-----SDWMKNLIS-----PD-----LGVVFLNVPENLKDPE----- 369
 QY 337 TPGPSLSQGVSVDEKLMPSAPVNTTIVADTSEQADTWLSERPK----- 382
 Db 370 ---PNIKMKRSIEEACFTFLQYN-----KLSMKPEPLFRSVGNTIEPVI 410
 QY 383 -----EIKVSKMEQKFRMLSDAPTQKESCKTSSNNNSV----- 417

Db 411 LFKQMGVGLKLEMYLVNPKVSSKEMQYF--MQQWTGNTKDKAEILILPNCQEVYDIPISYLS 468
 QY 418 -----SNTLAKMRIPNYQLSPTKLPISINKS-----DRASQOQNTSI--- 455
 Db 469 VSSLIVHPANPAEKIIRVLPFGNSTQYNIILEGLEKLLKHLDFLQPLATKDLTGQVPTP 528
 QY 456 -----RNYQPSTK-----KRDEENQEMSSCKSARIETSCSLLLEQTOP 495
 Db 529 PVKOVKLRADSRSLKPKATPKVASKSVRKSKEETPEVT--KTSQVE----- 575
 QY 496 ATPSLWNKKEOHLSENPEVDNSNNLFTDPLKSIKVNKASKSHAAEKLRSNKKREM-D 554
 Db 576 KTKPVESKEVLVKDKAPVKTESKPSV--TEKEVSKESQSPVKAIEVQKATESPKPKY 634
 QY 555 DVAIEDEV---LEQ-----LFKDTKP-----ELEIDVKVQ--KOEEDVNV 589
 Db 635 DKVVYKELIKTLEEKKEKPKKEVVKEDKTPKDKDEPKREEVYKKEIKKEIKERKEL 694
 QY 590 KRPMDIETNDTFSDEAVPESKISQENIEIGKREKEDSLWSAKEISNNKLOD---D 646
 Db 695 KKEVKKETPLKDA--KKEVKKEEKKEVKEKPEKKEIKKIS---KDIKKSTPQSDTKKP 749
 QY 647 SEMLPKLLLTPEPSLSLVKINSTSRNP--SGINDDYGLKNEKPKKVTYPG-----AGK 698
 Db 750 SALAPKVAKKEE-----STKKEPLAAGKLDKGKVKVVKKEGKTTEAAATAVGTAA 801
 QY 699 LPHIIGSDLIHARKNTELEEWLRQEME-----VONQHAKEESLADDL 743
 Db 802 TAAVVAAGTAASGPVKELEAERSLMSSPEDLTDKDFELKAEIDVAKDI 851
 RESULT 8
 BCK1 YEAST
 ID BCK1 YEAST STANDARD; PRT; 1478 AA.
 AC Q01389; P32894;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Serine/threonine protein kinase BCK1/SLK1/SSP31 (EC 2.7.-.-).
 GN BCK1 OR SLK1 OR SSP31 OR LAS3 OR YJL095W OR J0906.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID:4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92186847; PubMed=1545797;
 RA Costigan C., Gehring S., Snyder M.;
 RT "A synthetic lethal screen identifies SLK1, a novel protein kinase
 RT homolog implicated in yeast cell morphogenesis and cell growth.";
 RL Mol. Cell. Biol. 12:1162-1178(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92104496; PubMed=1840547;
 RA Irie K., Araki H., Oshima Y.;
 RT "A new protein kinase, SSP31, modulating the SMP3 gene-product
 RT involved in plasmid maintenance in Saccharomyces cerevisiae.";
 RL Gene 108:139-144(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EG123;
 RX MEDLINE=92107166; PubMed=1729597;
 RA Lee K.S., Levin D.E.;
 RT "Dominant mutations in a gene encoding a putative protein kinase
 RT (BCK1) bypass the requirement for a Saccharomyces cerevisiae protein
 RT kinase C homolog.";
 RL Mol. Cell. Biol. 12:172-182(1992).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=95176706; PubMed=7871887;
 RA Miosga T., Boles E., Schaaff-Gerstenschlaeger I., Schmitt S.,
 RA Zimmermann F.K.;

| | | | | | | | | | |
|----|----------|------|------|----------------------|----|----------|------|--|-----------------------------------|
| FT | DOMAIN | 1937 | 1989 | LAMININ EGF-LIKE 19. | FT | DISULFID | 1937 | 1951 | BY SIMILARITY. |
| FT | DOMAIN | 1990 | 2036 | LAMININ EGF-LIKE 20. | FT | DISULFID | 1939 | 1958 | BY SIMILARITY. |
| FT | DOMAIN | 2037 | 2083 | LAMININ EGF-LIKE 21. | FT | DISULFID | 1961 | 1970 | BY SIMILARITY. |
| FT | DOMAIN | 2084 | 2131 | LAMININ EGF-LIKE 22. | FT | DISULFID | 1973 | 1987 | BY SIMILARITY. |
| FT | DOMAIN | 2693 | 2884 | LAMININ G-LIKE 1. | FT | DISULFID | 1990 | 2000 | BY SIMILARITY. |
| FT | DOMAIN | 2896 | 3066 | LAMININ G-LIKE 2. | FT | DISULFID | 1992 | 2007 | BY SIMILARITY. |
| FT | DOMAIN | 3072 | 3235 | LAMININ G-LIKE 3. | FT | DISULFID | 2009 | 2018 | BY SIMILARITY. |
| FT | DOMAIN | 3310 | 3482 | LAMININ G-LIKE 4. | FT | DISULFID | 2021 | 2034 | BY SIMILARITY. |
| FT | DOMAIN | 3488 | 3669 | LAMININ G-LIKE 5. | FT | DISULFID | 2037 | 2048 | BY SIMILARITY. |
| FT | DISULFID | 298 | 307 | BY SIMILARITY. | FT | DISULFID | 2039 | 2055 | BY SIMILARITY. |
| FT | DISULFID | 300 | 320 | BY SIMILARITY. | FT | DISULFID | 2057 | 2066 | BY SIMILARITY. |
| FT | DISULFID | 322 | 331 | BY SIMILARITY. | FT | DISULFID | 2069 | 2081 | BY SIMILARITY. |
| FT | DISULFID | 334 | 354 | BY SIMILARITY. | FT | DISULFID | 2084 | 2096 | BY SIMILARITY. |
| FT | DISULFID | 357 | 366 | BY SIMILARITY. | FT | DISULFID | 2086 | 2103 | BY SIMILARITY. |
| FT | DISULFID | 359 | 391 | BY SIMILARITY. | FT | DISULFID | 2105 | 2114 | BY SIMILARITY. |
| FT | DISULFID | 394 | 403 | BY SIMILARITY. | FT | DISULFID | 2117 | 2129 | BY SIMILARITY. |
| FT | DISULFID | 406 | 424 | BY SIMILARITY. | FT | CARBOHYD | 121 | 121 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 427 | 438 | BY SIMILARITY. | FT | CARBOHYD | 140 | 140 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 429 | 445 | BY SIMILARITY. | FT | CARBOHYD | 249 | 249 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 447 | 456 | BY SIMILARITY. | FT | CARBOHYD | 351 | 351 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 459 | 469 | BY SIMILARITY. | FT | CARBOHYD | 477 | 477 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 472 | 484 | BY SIMILARITY. | FT | CARBOHYD | 511 | 511 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 474 | 491 | BY SIMILARITY. | FT | CARBOHYD | 530 | 530 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 493 | 502 | BY SIMILARITY. | FT | CARBOHYD | 634 | 634 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 505 | 516 | BY SIMILARITY. | FT | CARBOHYD | 761 | 761 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 519 | 531 | BY SIMILARITY. | FT | CARBOHYD | 1014 | 1014 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 521 | 538 | BY SIMILARITY. | FT | CARBOHYD | 1341 | 1341 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 540 | 549 | BY SIMILARITY. | FT | CARBOHYD | 1705 | 1705 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 552 | 561 | BY SIMILARITY. | FT | CARBOHYD | 1756 | 1756 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 564 | 576 | BY SIMILARITY. | FT | CARBOHYD | 1868 | 1868 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 566 | 583 | BY SIMILARITY. | FT | CARBOHYD | 1944 | 1944 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 585 | 594 | BY SIMILARITY. | FT | CARBOHYD | 1986 | 1986 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 597 | 607 | BY SIMILARITY. | FT | CARBOHYD | 2002 | 2002 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 610 | 622 | BY SIMILARITY. | FT | CARBOHYD | 2159 | 2159 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 612 | 629 | BY SIMILARITY. | FT | CARBOHYD | 2207 | 2207 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 631 | 640 | BY SIMILARITY. | FT | CARBOHYD | 2231 | 2231 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 643 | 653 | BY SIMILARITY. | FT | CARBOHYD | 2235 | 2235 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 656 | 668 | BY SIMILARITY. | FT | CARBOHYD | 2401 | 2401 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 658 | 674 | BY SIMILARITY. | FT | CARBOHYD | 2421 | 2421 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 676 | 685 | BY SIMILARITY. | FT | CARBOHYD | 2487 | 2487 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 688 | 698 | BY SIMILARITY. | FT | CARBOHYD | 2821 | 2821 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 701 | 715 | BY SIMILARITY. | FT | CARBOHYD | 3087 | 3087 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 703 | 724 | BY SIMILARITY. | FT | CARBOHYD | 3242 | 3242 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 726 | 735 | BY SIMILARITY. | FT | CARBOHYD | 3541 | 3541 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | DISULFID | 738 | 753 | BY SIMILARITY. | SQ | SEQUENCE | 3672 | AA; 404223.MW; 28E262DB5FF14BFA CRC64; | |
| FT | DISULFID | 756 | 770 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 758 | 777 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 779 | 788 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 791 | 806 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1415 | 1427 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1417 | 1434 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1436 | 1445 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1448 | 1458 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1461 | 1469 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1463 | 1476 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1478 | 1487 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1490 | 1503 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1506 | 1520 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1508 | 1527 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1529 | 1538 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1541 | 1551 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1554 | 1566 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1556 | 1573 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1575 | 1584 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1587 | 1602 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1830 | 1839 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1832 | 1846 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1849 | 1858 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1861 | 1877 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1880 | 1894 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1882 | 1905 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1907 | 1916 | BY SIMILARITY. | | | | | |
| FT | DISULFID | 1919 | 1934 | BY SIMILARITY. | | | | | |

Query Match 4.3%; Score 168.5; DB 1; Length 3672;
Best Local Similarity 20.4%; Pred. No. 0.6;
Matches 127; Conservative 105; Mismatches 201; Indels 189; Gaps 33;
QY 238 LSSAVFVGGEAR-----LITEENEENHF-----FLAPGTCVDTGI--- 275
Db 2160 ISSATIVGARLARNKKEFDINEITKMLNDEENSGNVFGDAQDILLNSTQIQKLVRTK 2219
QY 276 TNSQTLIPDCQK-----KWIQSIMDLQ--ROGLRPIPEAEIGLAVIFMTTKNYCDPQG 327
Db 2220 THSONSVSSAKNITLNGTEFLQEVMKRAQARSVSL--AEIALAI----- 2264
QY 328 HPSTGLKTTTGPSPSGSVDEKLMPSAPVNTTYY---VADTESEQADT--WDLSERPK 382
Db 2265 -----GSSKAVNVDPRLKAEETLMTLEAASADQYPERAQTVPGKLEETQK 2312
QY 383 EI-----KVKMEQKFR-----MLSDAPTKECKTSSNNNSVNTLAKMRI 426
Db 2313 KIQETEKLDKQKETFEAQKRAEELAYLSNAQQLKES--KSKADKSNNAKMLQJTKV 2371
QY 427 PN-----YQLSPTKLPISINKS--KDRASQ-----QQOTNSIRNYF 459
Db 2372 ENLVAAITDLDLVEAAKGEFQKLNVAIGNITENLKDREMTHAVITTLNETRNDVAEL 2431
QY 460 OPSTKKRDEEENQEMS--SCKSARIETSCSLLEQTOPATPSLWKNEQHLSENEPVDN 517

Db 2432 EAAKRVRRDEKSDVMQVLNNAKAHELHQAATTLRQT-----FDNN 2471
Qy 518 SDN-----NLF---TDT--DLKSVKNSASKSHAAEKLRSNKKRMDDDVAIEDEVLE 564
Db 2472 KDNTDQAVEAANAFSLNLTDLNKAQIDN-AYEALSAPAFASVQNRADKPPDETKE 2530
Qy 565 Q---LFKDTKPELEIDVKVQKEEDV-----NVRKR-----PRMDIETNDFTSDEAV 608
Db 2531 KIDALSRTVSODLKRETKLKKQLTSEKLRKRRAVRAKIPKYSKNPLDST-DEKV 2589
Qy 609 PESSKIQE-----NEI-GKKREL--KEDSLWSAKEISNNDKLODDSEMLPKK 653
Db 2590 QEVKLAEDIANTEETRAKISEIAGAEITERANGSAMGIRLARRNSVO-LNKLAP-- 2646
Qy 654 LLLTFEFLSL-VIKNSTSRNPNGINDYQGLKNFKFKKVTYYPGAGKPLPHIIGSDLIJAH 712
Db 2647 VIVSFEELKLSLSSAKVSDVSKYSQI---KEMIAVARDAANRIK-----LGAH 2695
Qy 713 ARKNTLEELWLRQEMEVQNOHA 734
Db 2696 FEKGSLLDNLTPQRVTRSAHA 2717
RESULT 11
TPR_HUMAN STANDARD; PRT; 2349 AA.
AC P12270;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleoprotein TPR.
GN TPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93064711; PubMed=1437155;
RA Mitchell P.J., Cooper C.S.;
RT "The human tpr gene encodes a protein of 2094 amino acids that has
RT extensive coiled-coil regions and an acidic C-terminal domain.";
RL Oncogene 7:2329-2333(1992).
RN [2]
RP REVISTONS, AND CHARACTERIZATION.
RX MEDLINE=95096166; PubMed=7798308;
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebl U., Gerace L.;
RT "tpr, a large coiled coil protein whose amino terminus is involved in
RT activation of oncogenic kinases, is localized to the cytoplasmic
RT surface of the nuclear pore complex.";
RL J. Cell Biol. 127:1515-1526(1994).
RN [3]
RP SEQUENCE OF 1-142 FROM N.A.
RX MEDLINE=98262257; PubMed=3387099;
RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
RT "tpr homologues activate met and raf.";
RL Oncogene 2:617-619(1988).
CC CC -!- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE
CC COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS
CC IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
CC TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
CC COMPONENTS, INCLUDING P62.
CC -!- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
CC BRAIN, LOWER LEVELS IN HEART, LIVER, AND KIDNEY.
CC -!- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
CC OR RAF GENES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X66397; CAA47021.1; -;
DR EMBL; Y00672; CAA68681.1; -;
DR PIR; S00928; S00928.
DR MM; 189940; -;
KW Coiled coil; Proto-oncogene; Chromosomal translocation;
KW Nuclear protein; Transport.
FT DOMAIN 78 360 COILED COIL (POTENTIAL).
FT DOMAIN 422 571 COILED COIL (POTENTIAL).
FT DOMAIN 575 628 COILED COIL (POTENTIAL).
FT DOMAIN 758 805 COILED COIL (POTENTIAL).
FT DOMAIN 834 869 COILED COIL (POTENTIAL).
FT DOMAIN 934 979 COILED COIL (POTENTIAL).
FT DOMAIN 1004 1064 COILED COIL (POTENTIAL).
FT DOMAIN 1138 1166 COILED COIL (POTENTIAL).
FT DOMAIN 1196 1241 COILED COIL (POTENTIAL).
FT DOMAIN 1262 1304 COILED COIL (POTENTIAL).
FT DOMAIN 1354 1434 COILED COIL (POTENTIAL).
FT DOMAIN 1476 1595 COILED COIL (POTENTIAL).
FT DOMAIN 527 530 POLY-SER.
FT DOMAIN 1833 1836 POLY-GLU.
FT DOMAIN 1957 1964 POLY-ASP.
FT DOMAIN 2295 2298 POLY-SER.
SQ SEQUENCE 2349 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;
Query Match 4.3%; Score 167.5; DB 1; Length 2349;
Best Local Similarity 22.1%; Pred. No. 0.39;
Matches 96; Conservative 85; Mismatches 183; Indels 71; Gaps 21;
Qy 333 LKTTTPGP---SLS-QGVSVDEKLMPSAVNTTIVADTESEQADTWLSPRKEIKVSK 388
Db 1174 VKGVQGLNVLSEBEGKSQEIILRFRIRREKIEFVQAQVESLRQ---RVEL 1230
Qy 389 MEQFRMLSDAPTVKESCKTSSNNNSMVSNTLAKMRIPNYQLSPTKLPISNKSDRASQ 448
Db 1231 LERELQELSDLSNAERKQVQVTAQTMAQHEELMKKTETMNVVMTNKM--LREEKERLEQ 1288
Qy 449 QQQNSTNRYFPSTKKRERD-----EENQEMSSCKSARIETSCSLLEQTOPATPSLWKN 503
Db 1289 DLQ-----QMOKVRKLELDILPLQEAANELSE-KSGMLQAEKKLLEE---DVKRWKA 1337
Qy 504 KEQHL-SENEPVDTSNDNLTFTDL--KSI-----VKSASKSHAA-----EKLRS 547
Db 1338 RNQHLVSOQKDPDTEEVYRKLLSEKEVHTKRIQQLTEELGRKAEIARSNASLTNNQNLIQ 1397
Qy 548 NKKREMDDDVAIEDEVLEQLFKDTKPE-LEIDVKVQKEEDVNVRRKPRMDIETNDFTSDE 606
Db 1398 SLKEDLNKVRTEKETIQ---KDLDAKIIDIQEKVKTITQVKKIGRRYKTYQVEELKAQODK 1454
Qy 607 AVPESSKISQENE-----IGKKRELKEDSLWSAKEISNNDKLODDSEMLPKK--LLTEF 659
Db 1455 VMETSQSSGDHQBQHVSVQEMQELKE-TLNQAE--TKSKLSLESQVENLQTLSEKETEFA 1511
Qy 660 RSL-----VIKNSTSRNPNGINDYQGLKNFK-----KFKKVTPYAGKPLPHIIGGS 706
Db 1512 RNLEQTVFSLQSELSRLRQDLQDRTQEEQLRQKITEKEKTRKAIWAASKIAHLGKLV 1571
Qy 707 DLIAHARKNTELEEE 721
Db 1572 DQL---TKENEELAQ 1583
RESULT 12
ATRX CAEEL STANDARD; PRT; 1359 AA.
AC Q9U7E0; Q02061;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)

RN [7] SEQUENCE OF 1410-1935 FROM N.A.
 RP MEDLINE-88299163; PubMed-2969919;
 RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;
 RT "Molecular cloning and characterization of human cardiac alpha- and
 RT beta-form myosin heavy chain complementary DNA clones. Regulation of
 RT expression during development and pressure overload in human
 RT atrium.";
 RL J. Clin. Invest. 82:524-531(1988).
 RN [8]
 RN [8] SEQUENCE OF 785-1935 FROM N.A.
 RP TISSUE-Skeletal muscle;
 RC MEDLINE-90235862; PubMed-1691980;
 RX Bober E., Buchberger-Seid A., Braun T., Singh S., Goedde H.W.,
 RA Arnold H.H.;
 RT "Identification of three developmentally controlled isoforms of human
 RT myosin heavy chains.";
 RL Eur. J. Biochem. 189:55-65(1990).
 RN [9]
 RN [9] SEQUENCE OF 1393-1935 FROM N.A.
 RP MEDLINE-87192738; PubMed-3032769;
 RX Jandreski M.A., Liew C.-C.;
 RA "Construction of a human ventricular cDNA library and
 RT characterization of a beta myosin heavy chain cDNA clone.";
 RL Hum. Genet. 76:47-53(1987).
 RN [10]
 RN [10] REVIEW ON VARIANTS.
 RP MEDLINE-96039076; PubMed-8533830;
 RX Arai S., Matsuo K., Hirayama K., Sukurai H., Tamura M., Ozawa T.,
 RA Kimura M., Imanura S.-I., Furutani Y., Joh-O K., Kawana M., Takao A.,
 RA Hosoda S., Momma K.;
 RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in
 RT hypertrophic cardiomyopathy.";
 RL Am. J. Med. Genet. 58:267-276(1995).
 RN [11]
 RN [11] VARIANTS CMH1 GLU-256 AND ARG-741.
 RP MEDLINE-93248216; PubMed-8483915;
 RX Fananapazir L., Dalakas M.C., Cyran F., Cohn G., Epstein N.D.;
 RA "Missense mutations in the beta-myosin heavy-chain gene cause central
 RT core disease in hypertrophic cardiomyopathy.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:3993-3997(1993).
 RN [12]
 RN [12] VARIANTS CMH1 GLN-403.
 RP MEDLINE-90367131; PubMed-1975517;
 RX Geisterfer-Lowrance A.A.T., Kass S., Tanigawa G., Vosberg H.-P.,
 RA McKenna W., Seidman C.E., Seidman J.G.;
 RT "A molecular basis for familial hypertrophic cardiomyopathy: a beta
 RT cardiac myosin heavy chain gene missense mutation.";
 RL Cell 62:999-1006(1990).
 RN [13]
 RN [13] VARIANTS CMH1 GLN-249; GLN-403; CYS-453 AND MET-606.
 RP MEDLINE-92204193; PubMed-1552912;
 RX Watkins H., Rosenzweig A., Hwang D.S., Levi T., McKenna W.,
 RA Seidmann C.E., Seidmann J.G.;
 RT "Characteristics and prognostic implications of myosin missense
 RT mutations in familial hypertrophic cardiomyopathy.";
 RL New Engl. J. Med. 326:1108-1114(1992).
 RN [14]
 RN [14] VARIANTS CMH1 GLN-403; CYS-453; ARG-584 AND MET-606.
 RP MEDLINE-94070863; PubMed-8250038;
 RX Watkins H., Thierfelder L., Anan R., Jarcho J., Matsumori A.,
 RA McKenna W., Seidman J.G., Seidman C.E.;
 RT "Independent origin of identical beta cardiac myosin heavy-chain
 RT mutations in hypertrophic cardiomyopathy.";
 RL Am. J. Hum. Genet. 53:1180-1185(1993).
 RN [15]
 RN [15] VARIANTS CMH1 GLN-403 AND VAL-908.
 RP MEDLINE-92346810; PubMed-1638703;
 RX Epstein N.D., Cohn G.M., Cyran F., Fananapazir L.;
 RA "Differences in clinical expression of hypertrophic cardiomyopathy
 RT associated with two distinct mutations in the beta-myosin heavy chain
 RT gene. A 908Leu-->Val mutation and a 403Arg-->Gln mutation.";
 RL Circulation 86:345-352(1992).

RN [16]
 RN [16] VARIANTS CMH1 LEU-403 AND TRP-403.
 RP MEDLINE-94075629; PubMed-8254035;
 RX Dausse E., Komajda M., Fetter L., Dubourg O., Dufour C., Carrier L.,
 RA Wisniewsky C., Berovic J., Hengstenberg C., Al-Mahdawi S.;
 RT "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and
 RT identification of a hot spot for mutations in the beta-myosin heavy
 RT chain gene.";
 RL J. Clin. Invest. 92:2807-2813(1993).
 RN [17]
 RN [17] VARIANTS CMH1 TRP-403.
 RP MEDLINE-94093568; PubMed-8268932;
 RX Moolman J.C., Brink P.A., Corfield V.A.;
 RT "Identification of a new missense mutation at Arg403, a CpG mutation
 RT hotspot, in exon 13 of the beta-myosin heavy chain gene in
 RT hypertrophic cardiomyopathy.";
 RL Hum. Mol. Genet. 2:1731-1732(1993).
 RN [18]
 RN [18] VARIANTS CMH1 ASN-615.
 RP MEDLINE-93038688; PubMed-1417858;
 RX Nishi H., Kimura A., Harada H., Toshima H., Sasazuki T.;
 RA "Novel missense mutation in cardiac beta-myosin heavy chain gene
 RT found in a Japanese patient with hypertrophic cardiomyopathy.";
 RL Biochem. Biophys. Res. Commun. 188:379-387(1992).
 RN [19]
 RN [19] VARIANTS CMH1 GLY-778.
 RP MEDLINE-93343938; PubMed-8343162;
 RX Harada H., Kimura A., Nishi H., Sasazuki T., Toshima H.;
 RA "A missense mutation of cardiac beta-myosin heavy chain gene linked
 RT to familial hypertrophic cardiomyopathy in affected Japanese
 RT families.";
 RL Biochem. Biophys. Res. Commun. 194:791-798(1993).
 RN [20]
 RN [20] VARIANTS CMH1 VAL-908.
 RP MEDLINE-93168485; PubMed-8435239;
 RX Al-Mahdawi S., Chamberlain S., Cleland J., Nihoyannopoulos P.,
 RA Gilligan D., French J., Choudhury L., Williamson R., Oakley C.;
 RT "Identification of a mutation in the beta cardiac myosin heavy chain
 RT gene in a family with hypertrophic cardiomyopathy.";
 RL Br. Heart J. 69:136-141(1993).
 RN [21]
 RN [21] VARIANTS CMH1 TRP-719.
 RP MEDLINE-95179132; PubMed-7874131;
 RX Greve G., Bachinski L., Friedman D.L., Czernuzewicz G., Anan R.,
 RA Towbin J., Seidman C.E., Roberts R.;
 RT "Isolation of a de novo mutant myocardial beta MHC protein in a
 RT pedigree with hypertrophic cardiomyopathy.";
 RL Hum. Mol. Genet. 3:2073-2075(1994).
 RN [22]
 RN [22] VARIANTS CMH1 CYS-513; ARG-716 AND TRP-719.
 RP MEDLINE-94110336; PubMed-8282798;
 RX Anan R., Greve G., Thierfelder L., Watkins H., McKenna W., Solomon S.,
 RA Vecchio C., Shono H., Nakao S., Tanaka H., Mares A. Jr., Towbin J.A.,
 RA Spirito P., Roberts R., Seidman J.G., Seidman C.E.;
 RT "Prognostic implications of novel beta cardiac myosin heavy chain gene
 RT mutations that cause familial hypertrophic cardiomyopathy.";
 RL J. Clin. Invest. 93:280-285(1994).
 RN [23]
 RN [23] VARIANTS CMH1 THR-797.
 RP MEDLINE-96047159; PubMed-7581410;
 RX Moolman J.C., Brink P.A., Corfield V.A.;
 RT "Identification of a novel Ala797Thr mutation in exon 21 of the beta-
 RT myosin heavy chain gene in hypertrophic cardiomyopathy.";
 RL Hum. Mutat. 6:197-198(1995).
 RN [24]
 RN [24] VARIANTS CMH1 CYS-453.
 RP MEDLINE-96209901; PubMed-8655135;
 RX Ko Y.-L., Chen J.-J., Tang T.-K., Cheng J.-J., Lin S.-Y., Liou Y.-C.,
 RA Kuan P., Wu C.-W., Lien W.-P., Liew C.-C.;
 RT "Malignant familial hypertrophic cardiomyopathy in a family with a
 RT 453Arg-->Cys mutation in the beta-myosin heavy chain gene;
 RT coexistence of sudden death and end-stage heart failure.";
 RL Hum. Genet. 97:585-590(1996).

RESULT 14

QY 369 SEQADTWDLSEPKKEIKVSKMEQKFRMLSDQAPTVKESCTSSNNNSWVNTLAK----- 423
 Db 195 SKDSISVSIKHOKKRYKSDSKE---MDFESTSTKRYKSKNNNSDAVSETCSKNSNV 251
 QY 424 -----MRIPNVQLS-----PTKLPSINKSKDRASQOQOQNSIRNYFQPS 462
 Db 252 GLMVHILGESDAESWEPDMWLKNTSQNNKSKPTKDKAKKDKAGKSGDAESVDS----- 303
 QY 463 TKRREDEENQESSCKSARIETSCSLE-----QTOPATPSPMLWKNKEQHLSENPEPVDTN 517
 Db 304 -KDAKKDKKCATKDTKKGAKKDTSTDAESGDSKDKAKKGGKSKDKKDKAKDAASDAE 362
 QY 518 SDNNLFTDTDLKSIVKNSAKSHAAEKLNRNKKREMDVAIEVQLFKDKYKPELEID 577
 Db 363 SG-----DSKDAKDK--SKKKDKSKDKNKKDKAKKDAESTDAESGDSKDKAKSKG 413
 QY 578 VKYKQOESDVNVRKRPMDJETNDTSDEA--VPESKISQENIEGKRELKEDSL-WSA 634
 Db 414 KDKSKDKDK---KDKAKKDAESTDAESGDSKNAKKDKSKGKKDD--KKDKAKKDAVSTDA 468
 QY 635 KEISNND--KLQDDSEMLPKKLLITFRSLVIRKSTSRNPSGINDDYQGLKNEFKFKKVT 692
 Db 469 DSESEGDAKSKDKSKDKKDLKDKDKQKPKAMK---SKESTETESDWESKKVKRDKSKDT 525
 QY 693 YPGAGKLPHIIGG-SOLIAHARKNTELEWLRQEMVQNHAKESLADLFRYNDYLK 751
 Db 526 KTKAKKATESGSAESDVSSRYLKKT-----EMFKSDAESSESL-----FKPGSK 571
 QY 752 RR 753
 Db 572 KR 573

RESULT 15
 MYH6 RAT
 ID MYH6_RAT STANDARD; PRT; 1938 AA.
 AC P02563; Q63351;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain, cardiac muscle alpha isoform (MyHC-alpha).
 GN MYH6.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=90016822; PubMed=2798111;
 RA Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
 RT "Complete nucleotide sequence of full length cDNA for rat alpha
 RT cardiac myosin heavy chain.";
 RL Nucleic Acids Res. 17:7527-7528(1989).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=90133919; PubMed=2614840;
 RA McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
 RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences.
 RT Comparisons suggest a molecular basis for functional differences.";
 RL J. Mol. Biol. 210:665-671(1989).
 RN [3]
 RP SEQUENCE OF 1-167 FROM N.A.
 RX MEDLINE=84194059; PubMed=6585819;
 RA Mahdavi V., Chambers A.P., Nadal-Ginard B.;
 RT "Cardiac alpha- and beta-myosin heavy chain genes are organized in
 RT tandem.";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:2626-2630(1984).
 RN [4]
 RP SEQUENCE OF 1512-1938 FROM N.A.
 RX MEDLINE=82220036; PubMed=7045682;
 RA Mahdavi V., Periasamy M., Nadal-Ginard B.;
 RT "Molecular characterization of two myosin heavy chain genes expressed

in the adult heart.";
 RL Nature 297:659-664(1982).
 RN [5]
 RP SEQUENCE OF 1872-1938 FROM N.A.
 RC STRAIN=Wistar; Tissue=Heart;
 RX MEDLINE=85179510; PubMed=6241892;
 RA Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;
 RT "Cardiac myosin heavy chain isozymic transitions during development
 RT and under pathological conditions are regulated at the level of mRNA
 RT availability.";
 RL Eur. Heart J. 5:181-191(1984).
 CC -!- FUNCTION: MUSCLE CONTRACTION.
 CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
 CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X15938; CAA34064.1; -
 CC EMBL: K01464; AAA41648.1; -
 CC EMBL: J00751; AAA41653.1; -
 CC EMBL: M32697; AAA41658.1; -
 CC PIR: S06005; S06005.
 CC PIR: A02988; A02988.
 CC HSSP: P08799; IIMD.
 CC InterPro: IPR000048; IQ.
 CC InterPro: IPR004009; Myosin_N.
 CC InterPro: IPR002928; Myosin_tail.
 CC InterPro: IPR001609; myosin_head.
 CC Pfam: PF00612; IQ; 2.
 CC Pfam: PF00063; myosin_head; 1.
 CC Pfam: PF02736; Myosin_N; 1.
 CC Pfam: PF01576; Myosin_tail; 1.
 CC PRINTS: PR00193; MYOSINHEAVY.
 CC ProDom: PD000355; myosin_head; 1.
 CC SMART: SM00015; IQ; 1.
 CC SMART: SM00242; MYSC; 1.
 CC PROSITE: PS50096; IQ; 1.
 CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 CC ATP-binding; Methylation; Alkylation; Multigene family;
 CC Calmodulin-binding.
 FT DOMAIN 1 781 MYOSIN HEAD-LIKE.
 FT DOMAIN 782 811 IQ.
 FT DOMAIN 842 1938 COILED COIL (POTENTIAL).
 FT NP_BIND 177 184 ATP.
 FT DOMAIN 656 678 ACTIN-BINDING.
 FT DOMAIN 758 772 ACTIN-BINDING.
 FT DOMAIN 789 806 CALMODULIN-BINDING (BY SIMILARITY).
 FT DOMAIN 815 832 CALMODULIN-BINDING (BY SIMILARITY).
 FT MOD_RES 128 128 METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 696 696 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 706 706 ALKYLATION (SH-2) (POTENTIAL).
 FT CONFLICT 13 13 R -> AP (IN REF. 3).

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:31:20 ; Search time 84.59 seconds
(without alignments)
12785.487 Million cell updates/sec

Title: US-09-837-602-1

Perfect score: 4403
Sequence: 1 ttccgacagcggcggttg.....accggtggagctcagct 4403

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| C 1 | 79.8 | 1.8 | 7218 | 1 | US-08-232-463-14 |
| C 2 | 65 | 1.5 | 19124 | 2 | US-08-487-826B-13 |
| C 3 | 59 | 1.3 | 7218 | 1 | US-08-232-463-14 |
| C 4 | 59 | 1.3 | 19124 | 2 | US-08-487-826B-13 |
| C 5 | 57.2 | 1.3 | 665 | 2 | US-08-883-795A-36 |
| C 6 | 55.6 | 1.3 | 3023 | 4 | US-09-308-022-4 |
| C 7 | 54.8 | 1.2 | 406 | 3 | US-08-928-799A-1 |
| C 8 | 54.8 | 1.2 | 545 | 4 | US-09-227-357-125 |
| C 9 | 54.8 | 1.2 | 752 | 4 | US-08-376-259-108 |
| C 10 | 54.8 | 1.2 | 849 | 3 | US-08-646-538-14 |
| C 11 | 54.8 | 1.2 | 849 | 4 | US-09-503-222-14 |
| C 12 | 54.8 | 1.2 | 912 | 2 | US-09-090-567-3 |
| C 13 | 54.8 | 1.2 | 1378 | 1 | US-08-075-533-20 |
| C 14 | 54.8 | 1.2 | 1378 | 2 | US-08-948-176-20 |
| C 15 | 54.8 | 1.2 | 1378 | 5 | PCT-US91-09160-20 |
| C 16 | 54.8 | 1.2 | 1556 | 4 | US-09-043-937A-3 |
| C 17 | 54.8 | 1.2 | 1568 | 4 | US-09-043-937A-1 |
| C 18 | 54.8 | 1.2 | 1830 | 1 | US-08-343-733A-2 |
| C 19 | 54.8 | 1.2 | 1958 | 4 | US-09-215-221-9 |
| C 20 | 54.8 | 1.2 | 2082 | 4 | US-09-293-505-9 |
| C 21 | 54.8 | 1.2 | 2605 | 2 | US-08-680-395-4 |
| C 22 | 54.8 | 1.2 | 2775 | 1 | US-08-730-771-1 |
| C 23 | 54.8 | 1.2 | 2775 | 4 | US-09-060-208-1 |
| C 24 | 54.8 | 1.2 | 2961 | 4 | US-08-446-935-6 |
| C 25 | 54.8 | 1.2 | 3089 | 1 | US-08-472-934-5 |
| C 26 | 54.8 | 1.2 | 3089 | 2 | US-08-323-460A-5 |
| C 27 | 54.8 | 1.2 | 3089 | 2 | US-08-461-146C-5 |

| | | | | | |
|------|------|-----|------|---|------------------|
| C 28 | 54.8 | 1.2 | 3089 | 3 | US-08-461-145C-5 |
| C 29 | 54.8 | 1.2 | 3089 | 4 | US-08-628-829-9 |
| C 30 | 54.8 | 1.2 | 3328 | 4 | US-08-960-048-1 |
| C 31 | 54.8 | 1.2 | 3699 | 3 | US-08-846-538-6 |
| C 32 | 54.8 | 1.2 | 3699 | 4 | US-09-503-222-6 |
| C 33 | 54.8 | 1.2 | 3792 | 2 | US-08-992-334-1 |
| C 34 | 54.8 | 1.2 | 3792 | 3 | US-08-302-752-1 |
| C 35 | 54.8 | 1.2 | 5234 | 2 | US-08-992-334-2 |
| C 36 | 54.8 | 1.2 | 5234 | 3 | US-08-302-752-2 |
| C 37 | 54.8 | 1.2 | 6045 | 3 | US-08-675-566-18 |
| C 38 | 54.8 | 1.2 | 6244 | 3 | US-08-675-566-17 |
| C 39 | 54.8 | 1.2 | 6447 | 3 | US-08-675-566-16 |
| C 40 | 54.8 | 1.2 | 6578 | 3 | US-08-675-566-4 |
| C 41 | 54.8 | 1.2 | 6612 | 3 | US-08-675-566-15 |
| C 42 | 54.8 | 1.2 | 6722 | 2 | US-08-992-334-3 |
| C 43 | 54.8 | 1.2 | 6722 | 3 | US-08-302-752-3 |
| C 44 | 54.8 | 1.2 | 6958 | 3 | US-08-675-566-2 |
| C 45 | 54.8 | 1.2 | 6994 | 3 | US-08-675-566-1 |

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)883-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
CLONE: ptzgpt-Fls
US-08-232-463-14

Query Match 1.3%; Score 59; DB 1; Length 7218;
Best Local Similarity 10.1%; Pred. No. 0.00022;
Matches 45; Conservative 213; Mismatches 189; Indels 0; Gaps 0;

QY 3819 ttctgtacctctgtcttttcaaaaatttggtgtgtccagttattttcccttttaac 3878
Db 1083 YYYVYYY 1142
QY 3879 cmttcccaattcgggtgtgtagtggtttccatttgggttttaattgttatccct 3938
Db 1143 YYYVYYY 1202
QY 3939 gatagctataattgggtcagataaattctttacattcttagatgaagtctcttgcgg 3998
Db 1203 YYYVYYY 1262
QY 3999 atatacgtattgatattacacctagctgtgtgcttgactgtttcttcttattgtctttg 4058
Db 1263 YYYVYYY 1322
QY 4059 atgaatagaagttttaaattttgacaaggtcacaatttatttttctttgttgatt 4118
Db 1323 YYYVYYY 1382
QY 4119 ttctctcoaatttaacccaagatttcagatatctctctattataaaactttata 4178
Db 1383 YYYVYYY 1442
QY 4179 tttttatttggatctaacctgaattgatgtatgttgtgaattatgatcagggtt 4238
Db 1443 TTCTTCTATCTCTTAACTACTTGTGCATAGATAGTAATTAACAGTGTGCTACATGCCGT 1502
QY 4239 ctttttttccccatcaagaatccag 4265
Db 1503 TTTTGTAAACTGAATAGATCGCTCTAG 1529

RESULT 4
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5933827
; GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Welles, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEITICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 1.3%; Score 59; DB 2; Length 19124;
Best Local Similarity 44.3%; Pred. No. 0.00031;
Matches 239; Conservative 0; Mismatches 300; Indels 0; Gaps 0;

QY 1618 attcagatcacagatttaaaatctattgtgaaaaattctgccagtaaatctcatgctgc 1677
Db 15449 AATGAAAAAAGAAAAATGAATATAAAAAAATTTTATATAAAAAAATTTTATATAAAAAA 15508
QY 1678 agaaagctaaatcaaaataaaaaaagggaatggatgagtgtgcccataagaatgaagt 1737
Db 15509 AAAAAAGGAGAAAAAATTTTATAAAAAATATAAAAAATTAATAAAAAATATAAAAAATTTTGAT 15568
QY 1738 attggaacagttattcaaggacacaaacccagagttagaaattgattgaaagtcaaaa 1797
Db 15569 AGAATAAAAAATGAAGAAGATTATCAAAAAAATTAATAAAAAATTTTATATAAAAAA 15628
QY 1798 acaggaggagatgtcaattgttagaaaaaggccaagatggatagataagaacaatgacac 1857
Db 15629 AAATGATTATAAAAAAATATAAAAAAAGAAAAAAGAAAAAATTAATAAAAAAATTAATAAAAAA 15688
QY 1858 ttccagtgatgaagcagtagccagaagaatagcaaaatctcagaagaatgaaattgggaa 1917
Db 15689 ATATATATATCAAAAAAAGAAAAAATTAATAAAAAAATTAATAAAAAATTAATAAAAAATATATCA 15748
QY 1918 gaaacgtgaactcaaggagagactcactatggtcagctcaagaataatctcaacaatgacaa 1977
Db 15749 TAAAAATAAAAAAATTAATAAAAAATTTTAAAAAATTAATAAAAAATTAATAAAAAA 15808
QY 1978 acttcaggatgatgtgagatgcttccaaaaaagctgttattgactgaatttagatcact 2037

; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 545
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (16)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (41)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (42)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (87)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-125

Query Match 1.2%; Score 54.8; DB 4; Length 545;
Best Local Similarity 83.3%; Pred. No. 0.00076;
Matches 50; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

Qy 4344 catmarmmmrrrggattccactagtcttagagcgccgcccacccggtggagctccagct 4403
| |:::|||||
Db 176 CCTGCAGCCGGGGATCCACTAGTCTTAGAGCGCCGCCACCCGCGGTGGAGCTCCAGCT 117

RESULT 9
US-08-976-259-108
; Sequence 108, Application US/08976259
; Patent No. 6316609
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Choi, Gil H.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli
; Patent No. 6316609
; NUMBER OF SEQUENCES: 142
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,259
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,626 AND US 60/061,953
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CBM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 108:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 752 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-976-259-108

Query Match 1.2%; Score 54.8; DB 4; Length 752;
Best Local Similarity 83.3%; Pred. No. 0.00086;
Matches 50; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

Qy 4344 catmarmmmrrrggattccactagtcttagagcgccgcccacccggtggagctccagct 4403
| |:::|||||
Db 52 CCTGCAGCCGGGGATCCACTAGTCTTAGAGCGCCGCCACCCGCGGTGGAGCTCCAGCT 111

RESULT 10
US-08-646-538-14/c
; Sequence 14, Application US/08646538
; Patent No. 6027881
; GENERAL INFORMATION:
; APPLICANT: Pavlakis, George N.
; APPLICANT: Gaitanaris, George A.
; APPLICANT: Stauber, Roland H.
; APPLICANT: Vournakis, John N.
; TITLE OF INVENTION: Mutant Aequorea victoria Fluorescent
; TITLE OF INVENTION: Proteins Having Increased Cellular Fluorescence
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,538
; FILING DATE: No. 6027881 yet assigned
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 015280-249000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 849 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

Query Match 1.2%; Score 54.8; DB 4; Length 849;
 Best Local Similarity 83.3%; Pred. No. 0.0009;
 Matches 50; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: E. I. du Pont de Nemours and Company
 STREET: 1007 Market Street
 CITY: Wilmington
 STATE: Delaware
 COUNTRY: U.S.A
 ZIP: 19898
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/075,533
 FILING DATE:
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/631,264
 FILING DATE: 20-DEC-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Morrissey, Bruce W.
 REGISTRATION NUMBER: 30,663
 REFERENCE/DOCKET NUMBER: CR-8926-A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (302) 992-4927
 TELEFAX: (302) 892-7949
 TELEX: 835420
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1378 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Brassica napus
 U08-075-533-20

| | | | | |
|-----------------------|------------------|--|-----------|--------------|
| Query Match | 1.2% | Score 54.8; | DB 1; | Length 1378; |
| Best Local Similarity | 83.3%; | Pred. No. 0.0011; | | |
| Matches | 50; Conservative | 9; Mismatches | 1; Indels | 0; Gaps 0; |
| QY | 4344 | catmarmmmrrrggatccactagtcttagagcggcgccacgcggtggagctccagct | 4403 | |
| Dd | 111 | CCTGCAGCCCGGGGATTCCTACTAGTCTACAGCGCGCGCACCCGCGTGAGCTCCAGCT | 52 | |

```

RESULT 14
US-08-948-176-20/c
; Sequence 20, Application US/08948176
; Patent No. 5945585
; GENERAL INFORMATION:
; APPLICANT: HITZ, WILLIAM D.
; APPLICANT: YADAV, NARENDRA S.
; TITLE OF INVENTION: ACYL-ACP THIOESTERASES GENES
; TITLE OF INVENTION: AND THEIR USE IN ALTERING PLANT
; TITLE OF INVENTION: OIL COMPOSITION
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT
; SOFTWARE: MICROSOFT WORD VERSION 7.0
; CURRENT APPLICATION DATA:

```

—

```
, APPLICATION NUMBER: US/08/948,176
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/631,264
; FILING DATE: DECEMBER 20, 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: CHRISTENBURY, LYNNE M.
; REGISTRATION NUMBER: 30,971
; REFERENCE/DOCKET NUMBER: CR-8926-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-5481
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1378 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Brassica napus
US-08-948-176-20

Query Match          1.2%; Score 54.8; DB 2; Length 1378;
Best Local Similarity   83.3%; Pred. No. 0.0011;
Matches      50; Conservative    9; Mismatches     1; Indels     0; Gaps     0;

Qy 4344 catmarmmmrrrgatccactagtctagacgcggcgcccaccggtggagtccaacct 4403
| : ::::: ||||| |
Db 111 CCTGCAGCCGGGGAATCCTCACTAGTTCTTAGAGCGGCCGACCGCGGTGGACTCCAAGT 52
```

RESULT 15
PCT-US91-09160-20/c
; Sequence 20, Application PC/TUS9109160
; GENERAL INFORMATION:
; APPLICANT: Hitz, William D.
; APPLICANT: Yadav, Narendra S.
; TITLE OF INVENTION: Nucleotide Sequences of Soybean Acyl-ACP
; TITLE OF INVENTION: Thioesterase Genes
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/09160
; FILING DATE: 19911216
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/631,264
; FILING DATE: 20-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Morrissey, Bruce W.
; REGISTRATION NUMBER: 30,663
; REFERENCE/DOCKET NUMBER: CR-8926-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4927
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1378 base pairs

```
;
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Brassica napus
PCT-US91-09160-20
```

```
Query Match 1.2%; Score 54.8; DB 5; Length 1378;
Best Local Similarity 83.3%; Pred. No. 0.0011;
Matches 50; Conservative 9; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 4344 catrmarmrrrrggatccactagtctagagcgccgccacccggtggagctccagct 4403
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 111 CCTGCAGCCCGGGGATCCACTAGTTCTAGAGCGCGCCACCCGCGTGGAGCTCCAGCT 52
```

```
Search completed: August 14, 2002, 23:56:47
Job time: 8727 sec
```

